

Diffusion Adaptation Over Networks Under Imperfect Information Exchange and Non-Stationary Data

Xiaochuan Zhao, *Student Member, IEEE*, Sheng-Yuan Tu, *Student Member, IEEE*, and Ali H. Sayed, *Fellow, IEEE*

Abstract

Adaptive networks rely on in-network and collaborative processing among distributed agents to deliver enhanced performance in estimation and inference tasks. Information is exchanged among the nodes, usually over noisy links. The combination weights that are used by the nodes to fuse information from their neighbors play a critical role in influencing the adaptation and tracking abilities of the network. This paper first investigates the mean-square performance of general adaptive diffusion algorithms in the presence of various sources of imperfect information exchanges, quantization errors, and model non-stationarities. Among other results, the analysis reveals that link noise over the regression data modifies the dynamics of the network evolution in a distinct way, and leads to biased estimates in steady-state. The analysis also reveals how the network mean-square performance is dependent on the combination weights. We use these observations to show how the combination weights can be optimized and adapted. Simulation results illustrate the theoretical findings and match well with theory.

Index Terms

Diffusion adaptation, adaptive networks, imperfect information exchange, tracking behavior, diffusion LMS, combination weights, energy conservation.

The authors are with Department of Electrical Engineering, University of California, Los Angeles, CA 90095, Email: {xzhao, shinetu, sayed}@ee.ucla.edu.

This work was supported in part by NSF grants CCF-0942936 and CCF-1011918. A short version of this work was accepted for the IEEE International Conference on Communications (ICC), Ottawa, ON, Canada, June 2012 [1].

I. INTRODUCTION

AN adaptive network consists of a collection of agents that are interconnected to each other and solve distributed estimation and inference problems in a collaborative manner. Two useful strategies that enable adaptation and learning over such networks in real-time are the incremental strategy [2]–[7] and the diffusion strategy [8]–[12]. Incremental strategies rely on the use of a Hamiltonian cycle, i.e., a cyclic path that covers all nodes in the network, which is generally difficult to enforce since determining a Hamiltonian cycle is an NP-hard problem. In addition, cyclic trajectories are not robust to node or link failure. In comparison, diffusion strategies are scalable, robust, and able to match well the performance of incremental networks. In adaptive diffusion implementations, information is processed locally at the nodes and then diffused in real-time across the network. Diffusion strategies were originally proposed in [8]–[10] and further extended and studied in [11]–[17]. They have been applied to model self-organized and complex behavior encountered in biological networks, such as fish schooling [18], bird flight formations [19], and bee swarming [20]. Diffusion strategies have also been applied to online learning of Gaussian mixture models [21], [22] and to general distributed optimization problems [23]. There have also been several useful works in the literature on distributed consensus-type strategies, with application to multi-agent formations and distributed processing [24]–[31]. The main difference between these works and the diffusion approach of [9], [11], [12] is the latter’s emphasis on the role of adaptation and learning over networks.

In the original diffusion least-mean-squares (LMS) strategy [9], [11], the weight estimates that are exchanged among the nodes can be subject to quantization errors and additive noise over the communication links. Studying the degradation in mean-square performance that results from these particular perturbations can be pursued, for both incremental and diffusion strategies, by extending the mean-square analysis already presented in [9], [11], in the same manner that the tracking analysis of conventional stand-alone adaptive filters was obtained from the counterpart results in the stationary case (as explained in [32, Ch. 21]). Useful results along these lines, which study the effect of link noise during the exchange of the weight estimates, already appear for the traditional diffusion algorithm in the works [33]–[36] and for consensus-based algorithms in [37], [38]. In this paper, our objective is to go beyond these earlier studies by taking into account additional effects, and by considering a more general algorithmic structure. The reason for this level of generality is because the analytical results will help reveal which noise sources influence the network performance more seriously, in what manner, and at what stage of the adaptation process. The results will suggest important remedies and mechanisms to adapt the combination weights

in real-time. Some of these insights are hard to get if one focuses solely on noise during the exchange of the weight estimates. The analysis will further show that noise during the exchange of the regression data plays a more critical role than other sources of imperfection: this particular noise alters the learning dynamics and modes of the network, and biases the weight estimates. Noises related to the exchange of other pieces of information do not alter the dynamics of the network but contribute to the deterioration of the network performance.

To arrive at these results, in this paper, we first consider a generalized analysis that applies to a broad class of diffusion adaptation strategies (see (5)–(7) further ahead; this class includes the original diffusion strategies (3) and (4) as two special cases). The analysis allows us to account for various sources of information noise over the communication links. We allow for noisy exchanges during *each* of the three processing steps of the adaptive diffusion algorithm (the two combination steps (5) and (7) and the adaptation step (6)). In this way, we are able to examine how the three sets of combination coefficients $\{a_{1,lk}, c_{lk}, a_{2,lk}\}$ in (5)–(7) influence the propagation of the noise signals through the network dynamics. Our results further reveal how the network mean-square-error performance is dependent on these combination weights. Following this line of reasoning, the analysis leads to algorithms (124) and (128) further ahead for choosing the combination coefficients to improve the steady-state network performance.

It should be noted that several combination rules, such as the Metropolis rule [39] and the maximum degree rule [40], were proposed previously in the literature — especially in the context of consensus-based iterations [40]–[42]. These schemes, however, usually suffer performance degradation in the presence of noisy information exchange since they ignore the network noise profile [15]. When the noise variance differs across the nodes, it becomes necessary to design combination rules that are aware of this variation as outlined further ahead in Section VI-B. Moreover, in a mobile network [18] where nodes are on the move and where neighborhoods evolve over time, it is even more critical to employ adaptive combination strategies that are able to track the variations in the noise profile in order to cope with such dynamic environments. This issue is taken up in Section VI-C.

A. Notation

We use lowercase letters to denote vectors, uppercase letters for matrices, plain letters for deterministic variables, and boldface letters for random variables. We also use $(\cdot)^*$ to denote conjugate transposition, $\text{Tr}(\cdot)$ for the trace of its matrix argument, $\rho(\cdot)$ for the spectral radius of its matrix argument, \otimes for the Kronecker product, and $\text{vec}(\cdot)$ for a vector formed by stacking the columns of its matrix argument.

We further use $\text{diag}\{\dots\}$ to denote a (block) diagonal matrix formed from its arguments, and $\text{col}\{\dots\}$ to denote a column vector formed by stacking its arguments on top of each other. All vectors in our treatment are column vectors, with the exception of the regression vectors, $\mathbf{u}_{k,i}$, and the associated noise signals, $\mathbf{v}_{lk,i}^{(u)}$, which are taken to be row vectors for convenience of presentation.

II. DIFFUSION ALGORITHMS WITH IMPERFECT INFORMATION EXCHANGE

We consider a connected network consisting of N nodes. Each node k collects scalar measurements $\mathbf{d}_k(i)$ and $1 \times M$ regression data vectors $\mathbf{u}_{k,i}$ over successive time instants $i \geq 0$. Note that we use parenthesis to refer to the time-dependence of scalar variables, as in $\mathbf{d}_k(i)$, and subscripts to refer to the time-dependence of vector variables, as in $\mathbf{u}_{k,i}$. The measurements across all nodes are assumed to be related to an unknown $M \times 1$ vector w^o via a linear regression model of the form [32]:

$$\mathbf{d}_k(i) = \mathbf{u}_{k,i} w^o + \mathbf{v}_k(i) \quad (1)$$

where $\mathbf{v}_k(i)$ denotes the measurement or model noise with zero mean and variance $\sigma_{v,k}^2$. The vector w^o in (1) denotes the parameter of interest, such as the parameters of some underlying physical phenomenon, the taps of a communication channel, or the location of food sources or predators. Such data models are also useful in studies on hybrid combinations of adaptive filters [43]–[47].

The nodes in the network would like to estimate w^o by solving the following minimization problem:

$$\underset{w}{\text{minimize}} \quad \sum_{k=1}^N \mathbb{E} |\mathbf{d}_k(i) - \mathbf{u}_{k,i} w|^2 \quad (2)$$

In previous works [9], [11], [13], we introduced and studied several distributed strategies of the diffusion type that allow nodes to cooperate with each other in order to solve problems of the form (2) in an adaptive manner. These diffusion strategies endow networks with adaptation and learning abilities, and enable information to diffuse through the network in real-time. We review the adaptive diffusion strategies below.

A. Diffusion Adaptation with Perfect Information Exchange

In [9], [11], two classes of diffusion algorithms were proposed. One class is the so-called Combine-then-Adapt (CTA) strategy:

$$\left\{ \begin{array}{l} \phi_{k,i-1} = \sum_{l \in \mathcal{N}_k} a_{1,lk} \mathbf{w}_{l,i-1} \\ \mathbf{w}_{k,i} = \phi_{k,i-1} + \mu_k \sum_{l \in \mathcal{N}_k} c_{lk} \mathbf{u}_{l,i}^* [\mathbf{d}_l(i) - \mathbf{u}_{l,i} \phi_{k,i-1}] \end{array} \right. \quad (3)$$

and the second class is the so-called Adapt-then-Combine (ATC) strategy:

$$\begin{cases} \boldsymbol{\psi}_{k,i} = \mathbf{w}_{k,i-1} + \mu_k \sum_{l \in \mathcal{N}_k} c_{lk} \mathbf{u}_{l,i}^* [\mathbf{d}_l(i) - \mathbf{u}_{l,i} \mathbf{w}_{k,i-1}] \\ \mathbf{w}_{k,i} = \sum_{l \in \mathcal{N}_k} a_{2,lk} \boldsymbol{\psi}_{l,i} \end{cases} \quad (4)$$

where the $\{\mu_k\}$ are small positive step-size parameters and the $\{a_{1,lk}, c_{lk}, a_{2,lk}\}$ are nonnegative entries of the $N \times N$ matrices $\{A_1, C, A_2\}$, respectively. The coefficients $\{a_{1,lk}, c_{lk}, a_{2,lk}\}$ are zero whenever node l is not connected to node k , i.e., $l \notin \mathcal{N}_k$, where \mathcal{N}_k denotes the neighborhood of node k . The two strategies (3) and (4) can be integrated into one broad class of diffusion adaptation [11]:

$$\boldsymbol{\phi}_{k,i-1} = \sum_{l \in \mathcal{N}_k} a_{1,lk} \mathbf{w}_{l,i-1} \quad (5)$$

$$\boldsymbol{\psi}_{k,i} = \boldsymbol{\phi}_{k,i-1} + \mu_k \sum_{l \in \mathcal{N}_k} c_{lk} \mathbf{u}_{l,i}^* [\mathbf{d}_l(i) - \mathbf{u}_{l,i} \boldsymbol{\phi}_{k,i-1}] \quad (6)$$

$$\mathbf{w}_{k,i} = \sum_{l \in \mathcal{N}_k} a_{2,lk} \boldsymbol{\psi}_{l,i} \quad (7)$$

Several diffusion strategies can be obtained as special cases of (5)–(7) through proper selection of the coefficients $\{a_{1,lk}, c_{lk}, a_{2,lk}\}$. For example, to recover the CTA strategy (3), we set $A_2 = I_N$, and to recover the ATC strategy (4), we set $A_1 = I_N$, where I_N denotes the $N \times N$ identity matrix. In the general diffusion strategy (5)–(7), each node k evaluates its estimate $\mathbf{w}_{k,i}$ at time i by relying solely on the data collected from its neighbors through steps (5) and (7) and on its local measurements through step (6). The matrices A_1 , A_2 , and C are required to be left or right-stochastic, i.e.,

$$A_1^T \mathbf{1}_N = \mathbf{1}_N, \quad A_2^T \mathbf{1}_N = \mathbf{1}_N, \quad C \mathbf{1}_N = \mathbf{1}_N \quad (8)$$

where $\mathbf{1}_N$ denotes the $N \times 1$ vector whose entries are all one. This means that each node performs a convex combination of the estimates received from its neighbors at every iteration i .

The mean-square performance and convergence properties of the diffusion algorithm (5)–(7) have already been studied in detail in [9], [11]. For the benefit of the analysis in the subsequent sections, we present below in (21) the recursion describing the evolution of the weight error vectors across the network. To do so, we introduce the error vectors:

$$\tilde{\boldsymbol{\phi}}_{k,i-1} \triangleq \mathbf{w}^o - \boldsymbol{\phi}_{k,i-1} \quad (9)$$

$$\tilde{\boldsymbol{\psi}}_{k,i} \triangleq \mathbf{w}^o - \boldsymbol{\psi}_{k,i} \quad (10)$$

$$\tilde{\mathbf{w}}_{k,i} \triangleq \mathbf{w}^o - \mathbf{w}_{k,i} \quad (11)$$

and substitute the linear model (1) into the adaptation step (6) to find that

$$\tilde{\psi}_{k,i} = (I_M - \mu_k \mathbf{R}_{k,i}) \tilde{\phi}_{k,i-1} - \mu_k \sum_{l \in \mathcal{N}_k} c_{lk} \mathbf{s}_{l,i} \quad (12)$$

where the $M \times M$ matrix $\mathbf{R}_{k,i}$ and the $M \times 1$ vector $\mathbf{s}_{k,i}$ are defined as:

$$\mathbf{R}_{k,i} \triangleq \sum_{l \in \mathcal{N}_k} c_{lk} \mathbf{u}_{l,i}^* \mathbf{u}_{l,i} \quad (13)$$

$$\mathbf{s}_{k,i} \triangleq \mathbf{u}_{k,i}^* \mathbf{v}_k(i) \quad (14)$$

We further collect the various quantities across all nodes in the network into the following block vectors and matrices:

$$\mathcal{R}_i \triangleq \text{diag} \{ \mathbf{R}_{1,i}, \dots, \mathbf{R}_{N,i} \} \quad (15)$$

$$\mathbf{s}_i \triangleq \text{col} \{ \mathbf{s}_{1,i}, \dots, \mathbf{s}_{N,i} \} \quad (16)$$

$$\mathcal{M} \triangleq \text{diag} \{ \mu_1 I_M, \dots, \mu_N I_M \} \quad (17)$$

$$\tilde{\phi}_i \triangleq \text{col} \{ \tilde{\phi}_{1,i}, \dots, \tilde{\phi}_{N,i} \} \quad (18)$$

$$\tilde{\psi}_i \triangleq \text{col} \{ \tilde{\psi}_{1,i}, \dots, \tilde{\psi}_{N,i} \} \quad (19)$$

$$\tilde{\mathbf{w}}_i \triangleq \text{col} \{ \tilde{\mathbf{w}}_{1,i}, \dots, \tilde{\mathbf{w}}_{N,i} \} \quad (20)$$

Then, from (5), (7), and (12), the recursion for the network error vector $\tilde{\mathbf{w}}_i$ is given by

$$\boxed{\tilde{\mathbf{w}}_i = \mathcal{A}_2^\top (I_{NM} - \mathcal{M} \mathcal{R}_i) \mathcal{A}_1^\top \tilde{\mathbf{w}}_{i-1} - \mathcal{A}_2^\top \mathcal{M} \mathcal{C}^\top \mathbf{s}_i} \quad (21)$$

where

$$\mathcal{A}_1 \triangleq A_1 \otimes I_M, \quad \mathcal{C} \triangleq C \otimes I_M, \quad \mathcal{A}_2 \triangleq A_2 \otimes I_M \quad (22)$$

B. Noisy Information Exchange

Each of the steps in (5)–(7) involves the sharing of information between node k and its neighbors. For example, in the first step (5), all neighbors of node k send their estimates $\mathbf{w}_{l,i-1}$ to node k . This transmission is generally subject to additive noise and possibly quantization errors. Likewise, steps (6) and (7) involve the sharing of other pieces of information with node k . These exchange steps can all be subject to perturbations (such as additive noise and quantization errors). One of the objectives of this work is to analyze the *aggregate* effect of these perturbations on general diffusion strategies of the type (5)–(7) and to propose choices for the combination weights in order to enhance the mean-square performance of the network in the presence of these disturbances.

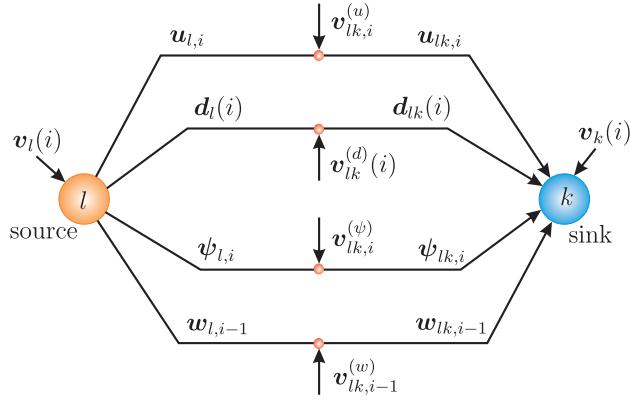


Fig. 1. Several additive noise sources perturb the exchange of information from node l to node k .

So let us examine what happens when information is exchanged over links with additive noise. We model the data received by node k from its neighbor l as

$$\mathbf{w}_{lk,i-1} \triangleq \mathbf{w}_{l,i-1} + \mathbf{v}_{lk,i-1}^{(w)} \quad (23)$$

$$\psi_{lk,i} \triangleq \psi_{l,i} + \mathbf{v}_{lk,i}^{(\psi)} \quad (24)$$

$$\mathbf{d}_{lk}(i) \triangleq \mathbf{d}_l(i) + \mathbf{v}_{lk,i}^{(d)}(i) \quad (25)$$

$$\mathbf{u}_{lk,i} \triangleq \mathbf{u}_{l,i} + \mathbf{v}_{lk,i}^{(u)} \quad (26)$$

where $\mathbf{v}_{lk,i-1}^{(w)}$ and $\mathbf{v}_{lk,i}^{(\psi)}$ are $M \times 1$ noise signals, $\mathbf{v}_{lk,i}^{(u)}$ is a $1 \times M$ noise signal, and $\mathbf{v}_{lk,i}^{(d)}(i)$ is a scalar noise signal (see Fig. 1). Observe further that in (23)–(26), we are including several sources of information exchange noise. In comparison, references [33]–[35] only considered the noise source $\mathbf{v}_{lk,i-1}^{(w)}$ in (23) and one set of combination coefficients $\{a_{1,lk}\}$; the other coefficients were set to $c_{lk} = a_{2,lk} = 0$ for $l \neq k$ and $c_{kk} = a_{2,kk} = 1$. In other words, these references only considered (23) and the following traditional CTA strategy without exchange of the data $\{\mathbf{d}_l(i), \mathbf{u}_{l,i}\}$ — compare with (3); note that the second step in (27) only uses $\{\mathbf{d}_k(i), \mathbf{u}_{k,i}\}$:

$$\begin{cases} \phi_{k,i-1} = \sum_{l \in \mathcal{N}_k} a_{1,lk} \mathbf{w}_{l,i-1} \\ \mathbf{w}_{k,i} = \phi_{k,i-1} + \mu_k \mathbf{u}_{k,i}^* [\mathbf{d}_k(i) - \mathbf{u}_{k,i} \phi_{k,i-1}] \end{cases} \quad (27)$$

The analysis that follows examines the aggregate effect of all four noise sources appearing in (23)–(26), in addition to the three sets of combination coefficients appearing in (5)–(7). We introduce the following assumption on the statistical properties of the measurement data and noise signals.

Assumption 2.1 (Statistical properties of the variables):

- 1) The regression data $\mathbf{u}_{k,i}$ are temporally white and spatially independent random variables with zero mean and covariance matrix $R_{u,k} \triangleq \mathbb{E} \mathbf{u}_{k,i}^* \mathbf{u}_{k,i} \geq 0$.
- 2) The noise signals $\mathbf{v}_k(i)$, $\mathbf{v}_{lk,i-1}^{(w)}$, $\mathbf{v}_{lk}^{(d)}(i)$, $\mathbf{v}_{lk,i}^{(u)}$, and $\mathbf{v}_{lk,i}^{(\psi)}$ are temporally white and spatially independent random variables with zero mean and (co)variances $\sigma_{v,k}^2$, $R_{v,lk}^{(w)}$, $\sigma_{v,lk}^2$, $R_{v,lk}^{(u)}$, and $R_{v,lk}^{(\psi)}$, respectively. In addition, $R_{v,lk}^{(w)}$, $\sigma_{v,lk}^2$, $R_{v,lk}^{(u)}$, and $R_{v,lk}^{(\psi)}$ are all zero if $l \notin \mathcal{N}_k$ or $l = k$.
- 3) The regression data $\{\mathbf{u}_{m,i_1}\}$, the model noise signals $\{\mathbf{v}_n(i_2)\}$, and the link noise signals $\{\mathbf{v}_{l_1 k_1, j_1}^{(w)}\}$, $\{\mathbf{v}_{l_2 k_2}^{(d)}(j_2)\}$, $\{\mathbf{v}_{l_3 k_3, j_3}^{(u)}\}$, and $\{\mathbf{v}_{l_4 k_4, j_4}^{(\psi)}\}$ are mutually-independent random variables for all $\{i_1, i_2, j_1, j_2, j_3, j_4\}$ and $\{m, n, l_1, l_2, l_3, l_4, k_1, k_2, k_3, k_4\}$. \blacksquare

Using the perturbed data (23)–(26), the diffusion algorithm (5)–(7) becomes

$$\phi_{k,i-1} = \sum_{l \in \mathcal{N}_k} a_{1,lk} \mathbf{w}_{lk,i-1} \quad (28)$$

$$\psi_{k,i} = \phi_{k,i-1} + \mu_k \sum_{l \in \mathcal{N}_k} c_{lk} \mathbf{u}_{lk,i}^* [\mathbf{d}_{lk}(i) - \mathbf{u}_{lk,i} \phi_{k,i-1}] \quad (29)$$

$$\mathbf{w}_{k,i} = \sum_{l \in \mathcal{N}_k} a_{2,lk} \psi_{lk,i} \quad (30)$$

where we continue to use the symbols $\{\phi_{k,i-1}, \psi_{k,i}, \mathbf{w}_{k,i}\}$ to avoid an explosion of notation. From (23) and (24), expressions (28)–(30) can be rewritten as

$$\phi_{k,i-1} = \sum_{l \in \mathcal{N}_k} a_{1,lk} \mathbf{w}_{l,i-1} + \mathbf{v}_{k,i-1}^{(w)} \quad (31)$$

$$\psi_{k,i} = \phi_{k,i-1} + \mu_k \sum_{l \in \mathcal{N}_k} c_{lk} \mathbf{u}_{lk,i}^* [\mathbf{d}_{lk}(i) - \mathbf{u}_{lk,i} \phi_{k,i-1}] \quad (32)$$

$$\mathbf{w}_{k,i} = \sum_{l \in \mathcal{N}_k} a_{2,lk} \psi_{l,i} + \mathbf{v}_{k,i}^{(\psi)} \quad (33)$$

where we are introducing the symbols $\mathbf{v}_{k,i-1}^{(w)}$ and $\mathbf{v}_{k,i}^{(\psi)}$ to denote the aggregate $M \times 1$ zero-mean noise signals defined over the neighborhood of node k :

$$\mathbf{v}_{k,i-1}^{(w)} \triangleq \sum_{l \in \mathcal{N}_k \setminus \{k\}} a_{1,lk} \mathbf{v}_{lk,i-1}^{(w)} \quad (34)$$

$$\mathbf{v}_{k,i}^{(\psi)} \triangleq \sum_{l \in \mathcal{N}_k \setminus \{k\}} a_{2,lk} \mathbf{v}_{lk,i}^{(\psi)} \quad (35)$$

with covariance matrices

$$R_{v,k}^{(w)} \triangleq \sum_{l \in \mathcal{N}_k \setminus \{k\}} a_{1,lk}^2 R_{v,lk}^{(w)} \quad (36)$$

$$R_{v,k}^{(\psi)} \triangleq \sum_{l \in \mathcal{N}_k \setminus \{k\}} a_{2,lk}^2 R_{v,lk}^{(\psi)} \quad (37)$$

It is worth noting that $R_{v,k}^{(w)}$ and $R_{v,k}^{(\psi)}$ depend on the combination coefficients $\{a_{1,lk}\}$ and $\{a_{2,lk}\}$, respectively. This property will be taken into account when optimizing over $\{a_{1,lk}\}$ and $\{a_{2,lk}\}$ in a later section. We further introduce the following scalar zero-mean noise signal:

$$\mathbf{v}_{lk}(i) \triangleq \mathbf{v}_l(i) + \mathbf{v}_{lk}^{(d)}(i) - \mathbf{v}_{lk,i}^{(u)} w^o \quad (38)$$

for $l \in \mathcal{N}_k \setminus \{k\}$, whose variance is

$$\sigma_{lk}^2 \triangleq \sigma_{v,l}^2 + \sigma_{v,lk}^2 + w^{o*} R_{v,lk}^{(u)} w^o \quad (39)$$

To unify the notation, we define $\mathbf{v}_{kk}(i) \triangleq \mathbf{v}_k(i)$. Then, from (1), (25), and (26), it is easy to verify that the noisy data $\{\mathbf{d}_{lk}(i), \mathbf{u}_{lk,i}\}$ are related via

$$\mathbf{d}_{lk}(i) = \mathbf{u}_{lk,i} w^o + \mathbf{v}_{lk}(i) \quad (40)$$

for $l \in \mathcal{N}_k$. Continuing with the adaptation step (32) and substituting (40), we get

$$\psi_{k,i} = \phi_{k,i-1} + \mu_k \sum_{l \in \mathcal{N}_k} c_{lk} \mathbf{u}_{lk,i}^* [\mathbf{u}_{lk,i} \tilde{\phi}_{k,i-1} + \mathbf{v}_{lk}(i)] \quad (41)$$

Then, we can derive the following error recursion for node k (compare with (12)):

$$\tilde{\psi}_{k,i} = (I_M - \mu_k \mathbf{R}'_{k,i}) \tilde{\phi}_{k,i-1} - \mu_k \mathbf{z}_{k,i} \quad (42)$$

where the $M \times M$ matrix $\mathbf{R}'_{k,i}$ and the $M \times 1$ vector $\mathbf{z}_{k,i}$ are defined as (compare with (13) and (14)):

$$\mathbf{R}'_{k,i} \triangleq \sum_{l \in \mathcal{N}_k} c_{lk} \mathbf{u}_{lk,i}^* \mathbf{u}_{lk,i} \quad (43)$$

$$\mathbf{z}_{k,i} \triangleq \sum_{l \in \mathcal{N}_k} c_{lk} \mathbf{u}_{lk,i}^* \mathbf{v}_{lk}(i) \quad (44)$$

We further introduce the block vectors and matrices:

$$\mathbf{R}'_i \triangleq \text{diag} \{ \mathbf{R}'_{1,i}, \dots, \mathbf{R}'_{N,i} \} \quad (45)$$

$$\mathbf{z}_i \triangleq \text{col} \{ \mathbf{z}_{1,i}, \dots, \mathbf{z}_{N,i} \} \quad (46)$$

$$\mathbf{v}_i^{(w)} \triangleq \text{col} \{ \mathbf{v}_{1,i}^{(w)}, \dots, \mathbf{v}_{N,i}^{(w)} \} \quad (47)$$

$$\mathbf{v}_i^{(\psi)} \triangleq \text{col} \{ \mathbf{v}_{1,i}^{(\psi)}, \dots, \mathbf{v}_{N,i}^{(\psi)} \} \quad (48)$$

and the corresponding covariance matrices for $\mathbf{v}_i^{(w)}$ and $\mathbf{v}_i^{(\psi)}$:

$$\mathcal{R}_v^{(w)} \triangleq \text{diag} \left\{ R_{v,1}^{(w)}, \dots, R_{v,N}^{(w)} \right\} \quad (49)$$

$$\mathcal{R}_v^{(\psi)} \triangleq \text{diag} \left\{ R_{v,1}^{(\psi)}, \dots, R_{v,N}^{(\psi)} \right\} \quad (50)$$

then, from (31), (33), and (42), we arrive at the following recursion for the network weight error vector in the presence of noisy information exchange:

$$\begin{aligned} \tilde{\mathbf{w}}_i &= \mathcal{A}_2^\top \tilde{\psi}_i - \mathbf{v}_i^{(\psi)} \\ &= \mathcal{A}_2^\top \left[(I_{NM} - \mathcal{M}\mathcal{R}'_i) \tilde{\phi}_{i-1} - \mathcal{M}\mathbf{z}_i \right] - \mathbf{v}_i^{(\psi)} \\ &= \mathcal{A}_2^\top \left[(I_{NM} - \mathcal{M}\mathcal{R}'_i) (\mathcal{A}_1^\top \tilde{\mathbf{w}}_{i-1} - \mathbf{v}_{i-1}^{(w)}) - \mathcal{M}\mathbf{z}_i \right] - \mathbf{v}_i^{(\psi)} \end{aligned} \quad (51)$$

That is,

$$\boxed{\tilde{\mathbf{w}}_i = \mathcal{A}_2^\top (I_{NM} - \mathcal{M}\mathcal{R}'_i) \mathcal{A}_1^\top \tilde{\mathbf{w}}_{i-1} - \mathcal{A}_2^\top (I_{NM} - \mathcal{M}\mathcal{R}'_i) \mathbf{v}_{i-1}^{(w)} - \mathcal{A}_2^\top \mathcal{M}\mathbf{z}_i - \mathbf{v}_i^{(\psi)}} \quad (52)$$

Compared to the previous error recursion (21), the noise terms in (52) consist of three parts:

- $\mathcal{A}_2^\top (I_{NM} - \mathcal{M}\mathcal{R}'_i) \mathbf{v}_{i-1}^{(w)}$ is contributed by the noise introduced at the information-exchange step (28) *before* adaptation.
- $\mathcal{A}_2^\top \mathcal{M}\mathbf{z}_i$ is contributed by the noise introduced at the adaptation step (29).
- $\mathbf{v}_i^{(\psi)}$ is contributed by the noise introduced at the information-exchange step (30) *after* adaptation.

III. CONVERGENCE IN THE MEAN WITH A BIAS

Given the weight error recursion (52), we are now ready to study the mean convergence condition for the diffusion strategy (28)–(30) in the presence of disturbances during information exchange under Assumption 2.1. Taking expectations of both sides of (52), we get

$$\mathbb{E}\tilde{\mathbf{w}}_i = \mathcal{B}\mathbb{E}\tilde{\mathbf{w}}_{i-1} - \mathcal{A}_2^\top (I_{NM} - \mathcal{M}\mathcal{R}') \cdot \mathbb{E}\mathbf{v}_{i-1}^{(w)} - \mathcal{A}_2^\top \mathcal{M} \cdot \mathbb{E}\mathbf{z}_i - \mathbb{E}\mathbf{v}_i^{(\psi)} \quad (53)$$

where

$$\mathcal{B} \triangleq \mathcal{A}_2^\top (I_{NM} - \mathcal{M}\mathcal{R}') \mathcal{A}_1^\top \quad (54)$$

$$\mathcal{R}' \triangleq \mathbb{E}\mathcal{R}'_i = \text{diag} \left\{ R'_1, \dots, R'_N \right\} \quad (55)$$

$$R'_k \triangleq \mathbb{E}\mathbf{R}'_{k,i} = \sum_{l \in \mathcal{N}_k} c_{lk} \left(R_{u,l} + R_{v,lk}^{(u)} \right) \quad (56)$$

From (34), (35), (47), and (48), it can be verified that

$$\mathbb{E}\mathbf{v}_{i-1}^{(w)} = \mathbb{E}\mathbf{v}_i^{(\psi)} = 0 \quad (57)$$

whereas, from (44) and Assumption 2.1, we get

$$\begin{aligned}\mathbb{E}\mathbf{z}_{k,i} &= \mathbb{E} \left[\sum_{l \in \mathcal{N}_k} c_{lk} (\mathbf{u}_{l,i} + \mathbf{v}_{lk,i}^{(u)})^* (\mathbf{v}_l(i) + \mathbf{v}_{lk}^{(d)}(i) - \mathbf{v}_{lk,i}^{(u)} w^o) \right] \\ &= - \left(\sum_{l \in \mathcal{N}_k} c_{lk} R_{v,lk}^{(u)} \right) w^o\end{aligned}\quad (58)$$

Let us define an $NM \times NM$ matrix $\mathcal{R}_{v,c}^{(u)}$ that collects all covariance matrices $\{R_{v,lk}^{(u)}\}$, $k, l = 1, \dots, N$, weighted by the corresponding combination coefficients $\{c_{lk}\}$, such that its (k, l) th $M \times M$ submatrix is $c_{lk} R_{v,lk}^{(u)}$. Note that $\mathcal{R}_{v,c}^{(u)}$ itself is *not* a covariance matrix because $c_{kk} R_{v,kk}^{(u)} = 0$ for all k . Then, from (46) and (58), we arrive at

$$z \triangleq \mathbb{E}\mathbf{z}_i = -\mathcal{R}_{v,c}^{(u)} (\mathbb{1}_N \otimes w^o) \quad (59)$$

Therefore, using (57) and (59), expression (53) becomes

$$\boxed{\mathbb{E}\tilde{\mathbf{w}}_i = \mathcal{B} \cdot \mathbb{E}\tilde{\mathbf{w}}_{i-1} - \mathcal{A}_2^\top \mathcal{M} z} \quad (60)$$

with a driving term due to the presence of z . This driving term would disappear from (60) if there were no noise during the exchange of the regression data. To guarantee convergence of (60), the coefficient matrix \mathcal{B} must be stable, i.e., $\rho(\mathcal{B}) < 1$. Since \mathcal{A}_1^\top and \mathcal{A}_2^\top are right-stochastic matrices, it can be shown that the matrix \mathcal{B} is stable whenever $I_{NM} - \mathcal{M}\mathcal{R}'$ itself is stable (see Appendix A). This fact leads to an upper bound on the step-sizes $\{\mu_k\}$ to guarantee the convergence of $\mathbb{E}\tilde{\mathbf{w}}_i$ to a steady-state value, namely, we must have

$$\boxed{\mu_k < \frac{2}{\lambda_{\max}(R'_k)}} \quad (61)$$

for $k = 1, 2, \dots, N$, where $\lambda_{\max}(\cdot)$ denotes the largest eigenvalue of its matrix argument. Note that the neighborhood covariance matrix R'_k in (56) is related to the combination weights $\{c_{lk}\}$. If we further assume that C is doubly-stochastic, i.e.,

$$C\mathbb{1}_N = \mathbb{1}_N, \quad C^\top \mathbb{1}_N = \mathbb{1}_N \quad (62)$$

then, by Jensen's inequality [48],

$$\begin{aligned}\lambda_{\max}(R'_k) &= \lambda_{\max} \left(\sum_{l \in \mathcal{N}_k} c_{lk} (R_{u,l} + R_{v,lk}^{(u)}) \right) \\ &\leq \sum_{l \in \mathcal{N}_k} c_{lk} \lambda_{\max} (R_{u,l} + R_{v,lk}^{(u)}) \\ &\leq \max_{l \in \mathcal{N}_k} \lambda_{\max} (R_{u,l} + R_{v,lk}^{(u)})\end{aligned}\quad (63)$$

since (i) $\lambda_{\max}(\cdot)$ coincides with the induced 2-norm for any positive semi-definite Hermitian matrix; (ii) matrix norms are convex functions of their arguments [49]; and (iii) by (62), $\{c_{lk}\}$ are convex combination coefficients. Thus, we obtain a sufficient condition for the convergence of (60) in lieu of (61):

$$\mu_k < \frac{2}{\max_{l \in \mathcal{N}_k} [\lambda_{\max}(R_{u,l} + R_{v,lk}^{(u)})]} \quad (64)$$

for $k = 1, 2, \dots, N$, where the upper bound for the step-size μ_k becomes independent of the combination weights $\{c_{lk}\}$. This bound can be determined solely from knowledge of the covariances of the regression data and the associated noise signals that are accessible to node k . It is worth noting that for traditional diffusion algorithms where information is perfectly exchanged, condition (64) reduces to

$$\mu_k < \frac{2}{\max_{l \in \mathcal{N}_k} [\lambda_{\max}(R_{u,l})]} \quad (65)$$

for $k = 1, 2, \dots, N$. Comparing (64) with (65), we see that the link noise $v_{lk,i}^{(u)}$ over regression data reduces the dynamic range of the step-sizes for mean stability. Now, under (61), and taking the limit of (60) as $i \rightarrow \infty$, we find that the mean error vector will converge to a steady-state value g :

$$g \triangleq \lim_{i \rightarrow \infty} \mathbb{E} \tilde{w}_i = -(I_{NM} - \mathcal{B})^{-1} \mathcal{A}_2^T \mathcal{M} z \quad (66)$$

IV. MEAN-SQUARE CONVERGENCE ANALYSIS

It is well-known that studying the mean-square convergence of a single adaptive filter is a challenging task, since adaptive filters are nonlinear, time-variant, and stochastic systems. When a network of adaptive nodes is considered, the complexity of the analysis is compounded because the nodes now influence each other's behavior. In order to make the performance analysis more tractable, we rely on the energy conservation approach [32], [50], which was used successfully in [9], [11] to study the mean-square performance of diffusion strategies under perfect information exchange conditions. That argument allows us to derive expressions for the mean-square-deviation (MSD) and the excess-mean-square-error (EMSE) of the network by analyzing how energy (measured in terms of error variances) flows through the nodes.

From recursion (52) and under Assumption 2.1, we can obtain the following weighted variance relation for the global error vector \tilde{w}_i :

$$\begin{aligned} \mathbb{E} \|\tilde{w}_i\|_{\Sigma}^2 &= \mathbb{E} \|\tilde{w}_{i-1}\|_{\Sigma'}^2 + \mathbb{E} \|\mathcal{A}_2^T \mathcal{M} z_i\|_{\Sigma}^2 - 2 \operatorname{Re} \{ \mathbb{E} [\mathcal{Z}_i^* \mathcal{M} \mathcal{A}_2 \Sigma \mathcal{A}_2^T (I_{NM} - \mathcal{M} \mathcal{R}'_i) \mathcal{A}_1^T \tilde{w}_{i-1}] \} \\ &\quad + \mathbb{E} \|\mathcal{A}_2^T (I_{NM} - \mathcal{M} \mathcal{R}'_i) v_{i-1}^{(w)}\|_{\Sigma}^2 + \mathbb{E} \|v_i^{(\psi)}\|_{\Sigma}^2 \end{aligned} \quad (67)$$

where Σ is an arbitrary $NM \times NM$ positive semi-definite Hermitian matrix that we are free to choose. Moreover, the notation $\|x\|_{\Sigma}^2$ stands for the quadratic term $x^* \Sigma x$. The weighting matrix Σ' in (67) can be expressed as

$$\Sigma' = \mathcal{B}^* \Sigma \mathcal{B} + O(\mathcal{M}^2) \quad (68)$$

where \mathcal{B} is given by (54) and $O(\mathcal{M}^2)$ denotes a term on the order of \mathcal{M}^2 . Evaluating the term $O(\mathcal{M}^2)$ requires knowledge of higher-order statistics of the regression data and link noises, which are not available under current assumptions. However, this term becomes negligible if we introduce a small step-size assumption.

Assumption 4.1 (Small step-sizes): The step-sizes are sufficiently small, i.e., $\mu_k \ll 1$, such that terms depending on higher-order powers of the step-sizes can be ignored. \blacksquare

Hence, in the sequel we use the approximation:

$$\Sigma' \approx \mathcal{B}^* \Sigma \mathcal{B} \quad (69)$$

Observe that on the right-hand side (RHS) of relation (67), only the first and third terms relate to the error vector $\tilde{\mathbf{w}}_{i-1}$. By Assumption 2.1, the error vector $\tilde{\mathbf{w}}_{i-1}$ is independent of \mathbf{z}_i and \mathcal{R}'_i . Thus, from (59), the third term on RHS of (67) can be expressed as

$$\begin{aligned} \text{Third term on RHS of (67)} &= -2 \operatorname{Re}\{\mathbb{E}[\mathbf{z}_i^* \mathcal{M} \mathcal{A}_2 \Sigma \mathcal{A}_2^T (I_{NM} - \mathcal{M} \mathcal{R}'_i) \mathcal{A}_1^T] \cdot \mathbb{E}\tilde{\mathbf{w}}_{i-1}\} \\ &= -2 \operatorname{Re}(z^* \mathcal{M} \mathcal{A}_2 \Sigma \mathcal{A}_2^T \mathcal{A}_1^T \cdot \mathbb{E}\tilde{\mathbf{w}}_{i-1}) + O(\mathcal{M}^2) \end{aligned} \quad (70)$$

Since we already showed in the previous section that $\mathbb{E}\tilde{\mathbf{w}}_i$ converges to a fixed bias g , quantity (70) will converge to a fixed value as well when $i \rightarrow \infty$. Moreover, under Assumption 2.1, the second, fourth, and fifth terms on RHS of relation (67) are all fixed values. Therefore, the convergence of relation (67) depends on the behavior of the first term $\mathbb{E}\|\tilde{\mathbf{w}}_{i-1}\|_{\Sigma'}^2$. Although the weighting matrix Σ' of $\tilde{\mathbf{w}}_{i-1}$ is different from the weighting matrix Σ of $\tilde{\mathbf{w}}_i$, it turns out that the entries of these two matrices are approximately related by a linear equation shown ahead in (72). Introduce the vector notation [32]:

$$\sigma = \operatorname{vec}(\Sigma), \quad \sigma' = \operatorname{vec}(\Sigma') \quad (71)$$

Then, by using the identity $\operatorname{vec}(ABC) = (C^T \otimes A) \cdot \operatorname{vec}(B)$, it can be verified from (69) that

$$\sigma' \approx \mathcal{F} \cdot \sigma \quad (72)$$

where the $N^2M^2 \times N^2M^2$ matrix \mathcal{F} is given by

$$\mathcal{F} \triangleq \mathcal{B}^T \otimes \mathcal{B}^* \quad (73)$$

To guarantee mean-square convergence of the algorithm, the step-sizes should be sufficiently small and selected to ensure that the matrix \mathcal{F} is stable [32], i.e., $\rho(\mathcal{F}) < 1$, which is equivalent to the earlier condition $\rho(\mathcal{B}) < 1$. Although more specific conditions for mean-square stability can be determined without Assumption 4.1 [32], it is sufficient for our purposes here to conclude that the diffusion strategy (28)–(30) is stable in the mean and mean-square senses if the step-sizes $\{\mu_k\}$ satisfy (61) or (64) and are sufficiently small.

V. STEADY-STATE PERFORMANCE ANALYSIS

The conclusion so far is that sufficiently small step-sizes ensure convergence of the diffusion strategy (28)–(30) in the mean and mean-square senses, even in the presence of exchange noises over the communication links. Let us now determine expressions for the error variances in steady-state. We start from the weighted variance relation (67). In view of (70), it shows that the error variance $\mathbb{E}\|\tilde{\mathbf{w}}_i\|_{\Sigma}^2$ depends on the mean error $\mathbb{E}\tilde{\mathbf{w}}_i$. We already determined the value of $\lim_{i \rightarrow \infty} \mathbb{E}\tilde{\mathbf{w}}_i$ in (66).

A. Steady-State Variance Relation

We continue to use the vector notation (71) and proceed to evaluate all the terms, except the first one, on RHS of (67) in the following. For the *second* term, it can be expressed as

$$\begin{aligned} \text{Second term on RHS of (67)} &= \text{Tr}(\mathcal{A}_2^T \mathcal{M} \mathcal{R}_z \mathcal{M} \mathcal{A}_2 \Sigma) \\ &= \left[\text{vec}(\mathcal{A}_2^T \mathcal{M} \mathcal{R}_z \mathcal{M} \mathcal{A}_2) \right]^* \sigma \end{aligned} \quad (74)$$

where we used the identity $\text{Tr}(W\Sigma) = [\text{vec}(W)]^* \sigma$ for any Hermitian matrix W , and \mathcal{R}_z denotes the autocorrelation matrix of \mathbf{z}_i . It is shown in Appendix B that \mathcal{R}_z is given by

$$\mathcal{R}_z \triangleq \mathbb{E} \mathbf{z}_i \mathbf{z}_i^* \approx \mathcal{C}^T \mathcal{S} \mathcal{C} + \mathcal{T} + z z^* \quad (75)$$

where \mathcal{C} is defined in (22), z is in (59), and $\{\mathcal{S}, \mathcal{T}\}$ are two $NM \times NM$ positive semi-definite block diagonal matrices:

$$\mathcal{S} \triangleq \text{diag} \left\{ \sigma_{v,1}^2 R_{u,1}, \dots, \sigma_{v,N}^2 R_{u,N} \right\} \quad (76)$$

$$\mathcal{T} \triangleq \text{diag} \{ T_1, \dots, T_N \} \quad (77)$$

$$T_k \triangleq \sum_{l \in \mathcal{N}_k} c_{lk}^2 \left[(\sigma_{v,l}^2 + \sigma_{v,lk}^2) R_{v,lk}^{(u)} + (\sigma_{v,lk}^2 + w^{o*} R_{v,lk}^{(u)} w^o) R_{u,l} \right] \quad (78)$$

From expression (70) and Assumption 4.1, the *third* term on RHS of (67) is given by

$$\begin{aligned} \text{Third term on RHS of (67)} &\approx -z^* \mathcal{M} \mathcal{A}_2 \Sigma \mathcal{A}_2^\top \mathcal{A}_1^\top (\mathbb{E} \tilde{\mathbf{w}}_{i-1}) - (\mathbb{E} \tilde{\mathbf{w}}_{i-1})^* \mathcal{A}_1 \mathcal{A}_2 \Sigma \mathcal{A}_2^\top \mathcal{M} z \\ &= -\text{Tr}\{[\mathcal{A}_2^\top \mathcal{A}_1^\top (\mathbb{E} \tilde{\mathbf{w}}_{i-1}) z^* \mathcal{M} \mathcal{A}_2 + \mathcal{A}_2^\top \mathcal{M} z (\mathbb{E} \tilde{\mathbf{w}}_{i-1})^* \mathcal{A}_1 \mathcal{A}_2] \Sigma\} \\ &= -\left[\text{vec} \left(\mathcal{A}_2^\top \mathcal{A}_1^\top (\mathbb{E} \tilde{\mathbf{w}}_{i-1}) z^* \mathcal{M} \mathcal{A}_2 + \mathcal{A}_2^\top \mathcal{M} z (\mathbb{E} \tilde{\mathbf{w}}_{i-1})^* \mathcal{A}_1 \mathcal{A}_2 \right) \right]^* \sigma \quad (79) \end{aligned}$$

Likewise, the *fourth* term on RHS of (67) is approximated by

$$\begin{aligned} \text{Fourth term on RHS of (67)} &= \left[\text{vec} \left(\mathbb{E} [\mathcal{A}_2^\top (I_{NM} - \mathcal{M} \mathcal{R}'_i) \mathcal{R}_v^{(w)} (I_{NM} - \mathcal{M} \mathcal{R}'_i) \mathcal{A}_2] \right) \right]^* \sigma \\ &\approx \left[\text{vec} (\mathcal{A}_2^\top \mathcal{R}_v^{(w)} \mathcal{A}_2) \right]^* \sigma \quad (80) \end{aligned}$$

where we are now ignoring terms on the order of \mathcal{M} and \mathcal{M}^2 . The *fifth* term on RHS of (67) is given by

$$\text{Fifth term on RHS of (67)} = \left[\text{vec} (\mathcal{R}_v^{(\psi)}) \right]^* \sigma \quad (81)$$

Let us introduce

$$\mathcal{R}_v \triangleq \mathcal{A}_2^\top \mathcal{R}_v^{(w)} \mathcal{A}_2 + \mathcal{R}_v^{(\psi)} + \mathcal{A}_2^\top \mathcal{M} (\mathcal{T} + z z^*) \mathcal{M} \mathcal{A}_2 \quad (82)$$

$$\begin{aligned} \mathcal{Y} &\triangleq -\mathcal{A}_2^\top \mathcal{A}_1^\top g z^* \mathcal{M} \mathcal{A}_2 \\ &= \mathcal{A}_2^\top \mathcal{A}_1^\top (I_{NM} - \mathcal{B})^{-1} \mathcal{A}_2^\top \mathcal{M} z z^* \mathcal{M} \mathcal{A}_2 \quad (83) \end{aligned}$$

At steady-state, as $i \rightarrow \infty$, by (66) and (74)–(83), the weighted variance relation (67) becomes

$$\lim_{i \rightarrow \infty} \mathbb{E} \|\tilde{\mathbf{w}}_i\|_\sigma^2 \approx \lim_{i \rightarrow \infty} \mathbb{E} \|\tilde{\mathbf{w}}_{i-1}\|_{\mathcal{F}\sigma}^2 + \left[\text{vec} (\mathcal{A}_2^\top \mathcal{M} \mathcal{C}^\top \mathcal{S} \mathcal{C} \mathcal{M} \mathcal{A}_2 + \mathcal{R}_v + \mathcal{Y} + \mathcal{Y}^*) \right]^* \sigma \quad (84)$$

where we are using the compact notation $\|x\|_\sigma^2$ to refer to $\|x\|_\Sigma^2$ — doing so allows us to represent Σ' by the more compact relation $\mathcal{F}\sigma$ on RHS of (84); we shall be using the weighting matrix Σ and its vector representation σ interchangeably for ease of notation (likewise, for Σ' and σ'). The steady-state weighted variance relation (84) can be rewritten as

$$\lim_{i \rightarrow \infty} \mathbb{E} \|\tilde{\mathbf{w}}_i\|_{(I_{N^2 M^2} - \mathcal{F})\sigma}^2 \approx \left[\text{vec} (\mathcal{A}_2^\top \mathcal{M} \mathcal{C}^\top \mathcal{S} \mathcal{C} \mathcal{M} \mathcal{A}_2 + \mathcal{R}_v + \mathcal{Y} + \mathcal{Y}^*) \right]^* \sigma \quad (85)$$

where the term $\mathcal{A}_2^\top \mathcal{M} \mathcal{C}^\top \mathcal{S} \mathcal{C} \mathcal{M} \mathcal{A}_2$ is contributed by the model noise $\{\mathbf{v}_k(i)\}$ while the remaining terms $\{\mathcal{R}_v, \mathcal{Y}\}$ are contributed by the link noises $\{\mathbf{v}_{lk,i-1}^{(w)}, \mathbf{v}_{lk}^{(d)}(i), \mathbf{v}_{lk,i}^{(u)}, \mathbf{v}_{lk,i}^{(\psi)}\}$. Recall that we are free to choose Σ and, hence, σ . Let $(I_{N^2 M^2} - \mathcal{F})\sigma = \text{vec}(\Omega)$, where Ω is another arbitrary positive semi-definite Hermitian matrix. Then, we arrive at the following theorem.

Theorem 5.1 (Steady-state weighted variance relation): Under Assumptions 2.1 and 4.1, for any positive semi-definite Hermitian matrix Ω , the steady-state weighted error variance relation of the diffusion strategy (28)–(30) is approximately given by

$$\lim_{i \rightarrow \infty} \mathbb{E} \|\tilde{\mathbf{w}}_i\|_{\Omega}^2 \approx \left[\text{vec}(\mathcal{A}_2^T \mathcal{M} \mathcal{C}^T \mathcal{S} \mathcal{C} \mathcal{M} \mathcal{A}_2 + \mathcal{R}_v + \mathcal{Y} + \mathcal{Y}^*) \right]^* (I_{N^2 M^2} - \mathcal{F})^{-1} \text{vec}(\Omega) \quad (86)$$

where \mathcal{S} is given in (76), \mathcal{R}_v in (82), \mathcal{Y} in (83), and \mathcal{F} in (73). \blacksquare

B. Network MSD and EMSE

Each subvector of $\tilde{\mathbf{w}}_i$ corresponds to the estimation error at a particular node, say, $\tilde{\mathbf{w}}_{k,i}$ for node k . The network MSD is defined as [32]:

$$\text{MSD} \triangleq \lim_{i \rightarrow \infty} \frac{1}{N} \sum_{k=1}^N \mathbb{E} \|\tilde{\mathbf{w}}_{k,i}\|^2 \quad (87)$$

Since we are free to choose Ω , we select it as $\Omega = I_{NM}/N$. Then, expression (86) gives

$$\text{MSD} \approx \frac{1}{N} \left[\text{vec}(\mathcal{A}_2^T \mathcal{M} \mathcal{C}^T \mathcal{S} \mathcal{C} \mathcal{M} \mathcal{A}_2 + \mathcal{R}_v + \mathcal{Y} + \mathcal{Y}^*) \right]^* (I_{N^2 M^2} - \mathcal{F})^{-1} \text{vec}(I_{NM}) \quad (88)$$

Similarly, if we instead select $\Omega = \mathcal{R}_u/N$, where

$$\mathcal{R}_u \triangleq \text{diag} \{R_{u,1}, \dots, R_{u,N}\} \quad (89)$$

then expression (86) would allow us to evaluate the network EMSE as:

$$\text{EMSE} \approx \frac{1}{N} \left[\text{vec}(\mathcal{A}_2^T \mathcal{M} \mathcal{C}^T \mathcal{S} \mathcal{C} \mathcal{M} \mathcal{A}_2 + \mathcal{R}_v + \mathcal{Y} + \mathcal{Y}^*) \right]^* (I_{N^2 M^2} - \mathcal{F})^{-1} \text{vec}(\mathcal{R}_u) \quad (90)$$

where, under Assumption 2.1, the network EMSE is given by

$$\begin{aligned} \text{EMSE} &\triangleq \lim_{i \rightarrow \infty} \frac{1}{N} \sum_{k=1}^N \mathbb{E} |\mathbf{u}_{k,i} \tilde{\mathbf{w}}_{k,i-1}|^2 \\ &= \lim_{i \rightarrow \infty} \frac{1}{N} \sum_{k=1}^N \mathbb{E} \|\tilde{\mathbf{w}}_{k,i}\|_{\mathcal{R}_u}^2 \end{aligned} \quad (91)$$

C. Simplifications when Regression Data are not Shared

We showed in the earlier sections that the link noise over regression data biases the weight estimators. In this section we examine how the results simplify when there is no sharing of regression data among the nodes.

Assumption 5.2 (No sharing of regression data): Nodes do not share regression data within neighborhoods, i.e., assume $\mathcal{C} = I_N$. \blacksquare

By Assumptions 4.1 and 5.2, matrices $\{\mathcal{B}, \mathcal{R}_v, \mathcal{Y}\}$ in (54), (82), and (83) become

$$\mathcal{B} = \mathcal{A}_2^\top (I_{NM} - \mathcal{M}\mathcal{R}_u) \mathcal{A}_1^\top \quad (92)$$

$$\mathcal{R}_v = \mathcal{A}_2^\top \mathcal{R}_v^{(w)} \mathcal{A}_2 + \mathcal{R}_v^{(\psi)} \quad (93)$$

$$\mathcal{Y} = 0 \quad (94)$$

where \mathcal{R}_u is given in (89). Then, the network MSD and EMSE expressions (88) and (90) simplify to:

$$\boxed{\text{MSD} \approx \frac{1}{N} \left[\text{vec}(\mathcal{A}_2^\top \mathcal{M}\mathcal{S}\mathcal{M}\mathcal{A}_2 + \mathcal{R}_v) \right]^* (I_{N^2M^2} - \mathcal{F})^{-1} \text{vec}(I_{NM})} \quad (95)$$

and

$$\boxed{\text{EMSE} \approx \frac{1}{N} \left[\text{vec}(\mathcal{A}_2^\top \mathcal{M}\mathcal{S}\mathcal{M}\mathcal{A}_2 + \mathcal{R}_v) \right]^* (I_{N^2M^2} - \mathcal{F})^{-1} \text{vec}(\mathcal{R}_u)} \quad (96)$$

D. Dependence of Performance on Combination Weights and Link Noise

Recalling that \mathcal{R}_v and \mathcal{F} are related to the combination matrices $\{\mathcal{A}_1, \mathcal{A}_2\}$, or, equivalently, $\{A_1, A_2\}$, results (95) and (96) express the network MSD and EMSE in terms of $\{A_1, A_2\}$. However, it is generally difficult to use these expressions to optimize over $\{A_1, A_2\}$ to reduce the impact of link noise. Instead, by substituting (73) into (95) and using the fact that \mathcal{F} is stable, we can arrive at another useful expression for the network MSD:

$$\begin{aligned} \text{MSD} &\approx \frac{1}{N} \left[\text{vec}(\mathcal{A}_2^\top \mathcal{M}\mathcal{S}\mathcal{M}\mathcal{A}_2 + \mathcal{R}_v) \right]^* \sum_{j=0}^{\infty} \mathcal{F}^j \text{vec}(I_{NM}) \\ &= \frac{1}{N} \left[\text{vec}(\mathcal{A}_2^\top \mathcal{M}\mathcal{S}\mathcal{M}\mathcal{A}_2 + \mathcal{R}_v) \right]^* \sum_{j=0}^{\infty} (\mathcal{B}^\top \otimes \mathcal{B}^*)^j \text{vec}(I_{NM}) \\ &= \frac{1}{N} \left[\text{vec}(\mathcal{A}_2^\top \mathcal{M}\mathcal{S}\mathcal{M}\mathcal{A}_2 + \mathcal{R}_v) \right]^* \sum_{j=0}^{\infty} \text{vec}(\mathcal{B}^{*j} \mathcal{B}^j) \end{aligned} \quad (97)$$

That is,

$$\boxed{\text{MSD} \approx \frac{1}{N} \sum_{j=0}^{\infty} \text{Tr} \left[\mathcal{B}^j (\mathcal{A}_2^\top \mathcal{M}\mathcal{S}\mathcal{M}\mathcal{A}_2 + \mathcal{R}_v) \mathcal{B}^{*j} \right]} \quad (98)$$

where \mathcal{B} is given in (92). Similarly, the network EMSE can be expressed as

$$\boxed{\text{EMSE} \approx \frac{1}{N} \sum_{j=0}^{\infty} \text{Tr} \left[\mathcal{B}^j (\mathcal{A}_2^\top \mathcal{M}\mathcal{S}\mathcal{M}\mathcal{A}_2 + \mathcal{R}_v) \mathcal{B}^{*j} \mathcal{R}_u \right]} \quad (99)$$

Expressions (98) and (99) reveal in an interesting way how the noise sources originating from any particular node end up influencing the overall network performance. Let us denote

$$\mathcal{B}_i \triangleq \mathcal{A}_2^\top (I_{NM} - \mathcal{M}\mathcal{R}'_i) \mathcal{A}_1^\top \quad (100)$$

$$\boldsymbol{\theta}_i \triangleq \mathcal{A}_2^\top (I_{NM} - \mathcal{M}\mathcal{R}'_i) \mathbf{v}_{i-1}^{(w)} + \mathcal{A}_2^\top \mathcal{M}\mathbf{z}_i + \mathbf{v}_i^{(\psi)} \quad (101)$$

The error recursion (52) can be rewritten as

$$\begin{aligned} \tilde{\mathbf{w}}_i &= \mathcal{B}_i \tilde{\mathbf{w}}_{i-1} - \boldsymbol{\theta}_i \\ &= \Phi_{0,i} \tilde{\mathbf{w}}_{-1} - \sum_{m=0}^i \Phi_{m+1,i} \boldsymbol{\theta}_m \end{aligned} \quad (102)$$

where

$$\Phi_{m,i} \triangleq \begin{cases} \mathcal{B}_i \mathcal{B}_{i-1} \dots \mathcal{B}_m, & i \geq m \\ I_{NM}, & i < m \end{cases} \quad (103)$$

Then,

$$\mathbb{E} \|\tilde{\mathbf{w}}_i\|^2 = \mathbb{E} \|\Phi_{0,i} \tilde{\mathbf{w}}_{-1}\|^2 + \mathbb{E} \left\| \sum_{m=0}^i \Phi_{m+1,i} \boldsymbol{\theta}_m \right\|^2 \quad (104)$$

Under Assumption 5.2, $\{\mathcal{B}_i, \boldsymbol{\theta}_i\}$ in (100) and (101) can be simplified as

$$\mathcal{B}_i = \mathcal{A}_2^\top (I_{NM} - \mathcal{M}\mathcal{R}_i) \mathcal{A}_1^\top \quad (105)$$

$$\boldsymbol{\theta}_i = \mathcal{A}_2^\top (I_{NM} - \mathcal{M}\mathcal{R}_i) \mathbf{v}_{i-1}^{(w)} + \mathcal{A}_2^\top \mathcal{M}\mathbf{s}_i + \mathbf{v}_i^{(\psi)} \quad (106)$$

where $\{\mathcal{R}_i, \mathbf{s}_i\}$ are given in (15) and (16). By Assumption 2.1, $\{\mathcal{B}_i, \boldsymbol{\theta}_i\}$ are temporally independent for different i and

$$\mathbb{E} \mathcal{B}_i = \mathcal{B}, \quad \mathbb{E} \boldsymbol{\theta}_i = 0 \quad (107)$$

where \mathcal{B} is given by (92). As $i \rightarrow \infty$, the first term on RHS of (104) becomes

$$\begin{aligned} \text{First term on RHS of (104)} &= \lim_{i \rightarrow \infty} \text{Tr} \left\{ \mathbb{E} [\Phi_{0,i} (\mathbb{E} \tilde{\mathbf{w}}_{-1} \tilde{\mathbf{w}}_{-1}^*) \Phi_{0,i}^*] \right\} \\ &\stackrel{(a)}{\approx} \lim_{i \rightarrow \infty} \text{Tr} \left[(\mathbb{E} \Phi_{0,i}) (\mathbb{E} \tilde{\mathbf{w}}_{-1} \tilde{\mathbf{w}}_{-1}^*) (\mathbb{E} \Phi_{0,i})^* \right] \\ &= \lim_{i \rightarrow \infty} \text{Tr} \left[\mathcal{B}^{i+1} (\mathbb{E} \tilde{\mathbf{w}}_{-1} \tilde{\mathbf{w}}_{-1}^*) \mathcal{B}^{(i+1)*} \right] \\ &\stackrel{(b)}{=} 0 \end{aligned} \quad (108)$$

where (a) is obtained by approximating the expectation of the product by the product of expectations and (b) is due to the stability of \mathcal{B} . Therefore, the steady-state value of (104) gives

$$\begin{aligned}
\lim_{i \rightarrow \infty} \mathbb{E} \|\tilde{\mathbf{w}}_i\|^2 &= \lim_{i \rightarrow \infty} \sum_{m=0}^i \mathbb{E} \|\Phi_{m+1,i} \boldsymbol{\theta}_m\|^2 \\
&\approx \lim_{i \rightarrow \infty} \sum_{m=0}^i \text{Tr} [(\mathbb{E} \Phi_{m+1,i}) (\mathbb{E} \boldsymbol{\theta}_m \boldsymbol{\theta}_m^*) (\mathbb{E} \Phi_{m+1,i})^*] \\
&\stackrel{(a)}{\approx} \lim_{i \rightarrow \infty} \sum_{m=0}^i \text{Tr} [\mathcal{B}^{i-m} (\mathcal{A}_2^\top \mathcal{M} \mathcal{S} \mathcal{M} \mathcal{A}_2 + \mathcal{R}_v) \mathcal{B}^{(i-m)*}] \\
&\stackrel{(b)}{=} \lim_{i \rightarrow \infty} \sum_{j=0}^i \text{Tr} [\mathcal{B}^j (\mathcal{A}_2^\top \mathcal{M} \mathcal{S} \mathcal{M} \mathcal{A}_2 + \mathcal{R}_v) \mathcal{B}^{j*}] \\
&= \sum_{j=0}^{\infty} \text{Tr} [\mathcal{B}^j (\mathcal{A}_2^\top \mathcal{M} \mathcal{S} \mathcal{M} \mathcal{A}_2 + \mathcal{R}_v) \mathcal{B}^{*j}]
\end{aligned} \tag{109}$$

where, by (93) and (106), (a) is due to

$$\begin{aligned}
\mathbb{E} \boldsymbol{\theta}_m \boldsymbol{\theta}_m^* &\approx \mathcal{A}_2^\top (I_{NM} - \mathcal{M} \mathcal{R}_u) \mathcal{R}_v^{(w)} (I_{NM} - \mathcal{M} \mathcal{R}_u) \mathcal{A}_2 + \mathcal{A}_2^\top \mathcal{M} \mathcal{S} \mathcal{M} \mathcal{A}_2 + \mathcal{R}_v^{(\psi)} \\
&\approx \mathcal{A}_2^\top \mathcal{M} \mathcal{S} \mathcal{M} \mathcal{A}_2 + \mathcal{A}_2^\top \mathcal{R}_v^{(w)} \mathcal{A}_2 + \mathcal{R}_v^{(\psi)} \\
&= \mathcal{A}_2^\top \mathcal{M} \mathcal{S} \mathcal{M} \mathcal{A}_2 + \mathcal{R}_v
\end{aligned} \tag{110}$$

and (b) is simply a change of variable: $j = i - m$. Since the j th term of the summation in (98) or (109) is contributed by the term $\mathbb{E} \boldsymbol{\theta}_{i-j} \boldsymbol{\theta}_{i-j}^*$, which consists of all the noise sources at time $i - j$, expression (98) shows how various sources of noises are involved and how they contribute to the network MSD.

VI. OPTIMIZING THE COMBINATION MATRICES

Before we optimize the combination matrices $\{A_1, A_2\}$, we first specialize the MSD expression (98) and the EMSE expression (99) for the ATC and CTA algorithms. For the ATC algorithm, we set $A_1 = I_N$ and $A_2 = A$, and for the CTA algorithm, we set $A_1 = A$ and $A_2 = I_N$. Let us denote

$$\mathcal{A} \triangleq A \otimes I_M \tag{111}$$

$$\mathcal{B}_{\text{atc}} \triangleq \mathcal{A}^\top (I_{NM} - \mathcal{M} \mathcal{R}_u) \tag{112}$$

$$\mathcal{B}_{\text{cta}} \triangleq (I_{NM} - \mathcal{M} \mathcal{R}_u) \mathcal{A}^\top \tag{113}$$

Then, we get

$$\text{MSD}_{\text{atc}} \approx \frac{1}{N} \sum_{j=0}^{\infty} \text{Tr} \left[\mathcal{B}_{\text{atc}}^j (\mathcal{A}^T \mathcal{M} \mathcal{S} \mathcal{M} \mathcal{A} + \mathcal{R}_v^{(\psi)}) \mathcal{B}_{\text{atc}}^{*j} \right] \quad (114)$$

$$\text{EMSE}_{\text{atc}} \approx \frac{1}{N} \sum_{j=0}^{\infty} \text{Tr} \left[\mathcal{B}_{\text{atc}}^j (\mathcal{A}^T \mathcal{M} \mathcal{S} \mathcal{M} \mathcal{A} + \mathcal{R}_v^{(\psi)}) \mathcal{B}_{\text{atc}}^{*j} \mathcal{R}_u \right] \quad (115)$$

and

$$\text{MSD}_{\text{cta}} \approx \frac{1}{N} \sum_{j=0}^{\infty} \text{Tr} \left[\mathcal{B}_{\text{cta}}^j (\mathcal{M} \mathcal{S} \mathcal{M} + \mathcal{R}_v^{(w)}) \mathcal{B}_{\text{cta}}^{*j} \right] \quad (116)$$

$$\text{EMSE}_{\text{cta}} \approx \frac{1}{N} \sum_{j=0}^{\infty} \text{Tr} \left[\mathcal{B}_{\text{cta}}^j (\mathcal{M} \mathcal{S} \mathcal{M} + \mathcal{R}_v^{(w)}) \mathcal{B}_{\text{cta}}^{*j} \mathcal{R}_u \right] \quad (117)$$

A. An Upper Bound for MSD

Minimizing the MSD expression (114) or the EMSE expression (115) for the ATC algorithm over left-stochastic matrices A is generally nontrivial. We pursue an approximate solution that relies on optimizing an upper bound and performs well in practice. Let us use $\|X\|_*$ to denote the nuclear norm (also known as the trace norm, or the Ky Fan n -norm) of matrix X [51], which is defined as the sum of the singular values of X . Therefore, $\|X\|_* = \|X^*\|_*$ for any X and $\|X\|_* = \text{Tr}(X)$ when X is Hermitian and positive semi-definite. Let us also denote $\|X\|_{b,\infty}$ as the block maximum norm of matrix X (see Appendix A).

Then,

$$\begin{aligned} \text{Tr} \left[\mathcal{B}_{\text{atc}}^j (\mathcal{A}^T \mathcal{M} \mathcal{S} \mathcal{M} \mathcal{A} + \mathcal{R}_v^{(\psi)}) \mathcal{B}_{\text{atc}}^{*j} \right] &= \left\| \mathcal{B}_{\text{atc}}^j (\mathcal{A}^T \mathcal{M} \mathcal{S} \mathcal{M} \mathcal{A} + \mathcal{R}_v^{(\psi)}) \mathcal{B}_{\text{atc}}^{*j} \right\|_* \\ &\leq \left\| \mathcal{B}_{\text{atc}}^j \right\|_* \cdot \left\| \mathcal{A}^T \mathcal{M} \mathcal{S} \mathcal{M} \mathcal{A} + \mathcal{R}_v^{(\psi)} \right\|_* \cdot \left\| \mathcal{B}_{\text{atc}}^{*j} \right\|_* \\ &\leq c^2 \cdot \left\| \mathcal{B}_{\text{atc}}^j \right\|_{b,\infty}^2 \cdot \text{Tr}(\mathcal{A}^T \mathcal{M} \mathcal{S} \mathcal{M} \mathcal{A} + \mathcal{R}_v^{(\psi)}) \\ &\leq c^2 \cdot \left\| \mathcal{B}_{\text{atc}} \right\|_{b,\infty}^{2j} \cdot \text{Tr}(\mathcal{A}^T \mathcal{M} \mathcal{S} \mathcal{M} \mathcal{A} + \mathcal{R}_v^{(\psi)}) \\ &\leq c^2 \cdot (\| \mathcal{A} \|_{b,\infty} \cdot \| I_{NM} - \mathcal{M} \mathcal{R}_u \|_{b,\infty})^{2j} \text{Tr}(\mathcal{A}^T \mathcal{M} \mathcal{S} \mathcal{M} \mathcal{A} + \mathcal{R}_v^{(\psi)}) \\ &= c^2 \cdot \rho(I_{NM} - \mathcal{M} \mathcal{R}_u)^{2j} \cdot \text{Tr}(\mathcal{A}^T \mathcal{M} \mathcal{S} \mathcal{M} \mathcal{A} + \mathcal{R}_v^{(\psi)}) \end{aligned} \quad (118)$$

where c is some positive scalar such that $\|X\|_* \leq c \|X\|_{b,\infty}$ because $\|X\|_*$ and $\|X\|_{b,\infty}$ are submultiplicative norms and all such norms are equivalent [49]. In the last step of (118) we used Lemmas A.4

and A.5 from Appendix A. Thus, we can upper bound the network MSD (114) by

$$\begin{aligned} \text{MSD}_{\text{atc}} &\leq \frac{1}{N} \sum_{j=0}^{\infty} c^2 \cdot \rho (I_{NM} - \mathcal{M}\mathcal{R}_u)^{2j} \text{Tr}(\mathcal{A}^T \mathcal{M} \mathcal{S} \mathcal{M} \mathcal{A} + \mathcal{R}_v^{(\psi)}) \\ &= \frac{c^2}{N} \cdot \frac{\text{Tr}(\mathcal{A}^T \mathcal{M} \mathcal{S} \mathcal{M} \mathcal{A} + \mathcal{R}_v^{(\psi)})}{1 - [\rho(I_{NM} - \mathcal{M}\mathcal{R}_u)]^2} \end{aligned} \quad (119)$$

where the combination matrix \mathcal{A} appears only in the numerator.

B. Minimizing the Upper Bound

The result (119) motivates us to consider instead the problem of minimizing the upper bound, namely,

$$\begin{aligned} \underset{\mathcal{A}}{\text{minimize}} \quad & \text{Tr}(\mathcal{A}^T \mathcal{M} \mathcal{S} \mathcal{M} \mathcal{A} + \mathcal{R}_v^{(\psi)}) \\ \text{subject to} \quad & \mathcal{A}^T \mathbf{1} = \mathbf{1}, \quad a_{lk} \geq 0, \quad a_{lk} = 0 \text{ if } l \notin \mathcal{N}_k \end{aligned} \quad (120)$$

Using (50) and (76), the cost function in (120) can be expressed as

$$\text{Tr}(\mathcal{A}^T \mathcal{M} \mathcal{S} \mathcal{M} \mathcal{A} + \mathcal{R}_v^{(\psi)}) = \sum_{k=1}^N \sum_{l \in \mathcal{N}_k} a_{lk}^2 \left[\mu_l^2 \sigma_{v,l}^2 \text{Tr}(R_{u,l}) + \text{Tr}(R_{v,lk}^{(\psi)}) \right] \quad (121)$$

Problem (120) can therefore be decoupled into N separate optimization problems of the form:

$$\begin{aligned} \underset{\{a_{lk}, l \in \mathcal{N}_k\}}{\text{minimize}} \quad & \sum_{l \in \mathcal{N}_k} a_{lk}^2 \left[\mu_l^2 \sigma_{v,l}^2 \text{Tr}(R_{u,l}) + \text{Tr}(R_{v,lk}^{(\psi)}) \right] \\ \text{subject to} \quad & \sum_{l \in \mathcal{N}_k} a_{lk} = 1, \quad a_{lk} \geq 0, \quad a_{lk} = 0 \text{ if } l \notin \mathcal{N}_k \end{aligned} \quad (122)$$

for $k = 1, \dots, N$. With each node $l \in \mathcal{N}_k$, we associate the following nonnegative *variance product* measure:

$$\gamma_{lk}^2 \triangleq \begin{cases} \mu_k^2 \sigma_{v,k}^2 \text{Tr}(R_{u,k}), & l = k \\ \mu_l^2 \sigma_{v,l}^2 \text{Tr}(R_{u,l}) + \text{Tr}(R_{v,lk}^{(\psi)}), & l \in \mathcal{N}_k \setminus \{k\} \end{cases} \quad (123)$$

This measure incorporates information about the link noise covariances $\{R_{v,lk}^{(\psi)}\}$. The solution of (122) is then given by

$$a_{lk} = \begin{cases} \frac{\gamma_{lk}^{-2}}{\sum_{m \in \mathcal{N}_k} \gamma_{mk}^{-2}}, & \text{if } l \in \mathcal{N}_k \\ 0, & \text{otherwise} \end{cases} \quad (\text{relative variance rule}) \quad (124)$$

We refer to this combination rule as the relative variance combination rule; it is an extension of the rule devised in [52] to the case of noisy information exchanges. In particular, the definition of the scalars $\{\gamma_{lk}^2\}$ in (123) is different and now depends on both subscripts l and k .

Minimizing the EMSE expression (115) for the ATC algorithm over left-stochastic matrices A can be pursued in a similar manner by noting that

$$\text{Tr} \left[\mathcal{B}_{\text{atc}}^j (\mathcal{A}^T \mathcal{M} \mathcal{S} \mathcal{M} \mathcal{A} + \mathcal{R}_v^{(\psi)}) \mathcal{B}_{\text{atc}}^{*j} \mathcal{R}_u \right] \leq c^2 [\rho(I_{NM} - \mathcal{M} \mathcal{R}_u)]^{2j} \text{Tr}(\mathcal{A}^T \mathcal{M} \mathcal{S} \mathcal{M} \mathcal{A} + \mathcal{R}_v^{(\psi)}) \text{Tr}(\mathcal{R}_u) \quad (125)$$

Thus, minimizing the upper bound of the network EMSE leads to the same solution (124). Using the same argument, we can also show that the same result minimizes the upper bound of the network MSD or EMSE for the CTA algorithm.

C. Adaptive Combination Rule

To apply the relative variance combination rule (124), each node k needs to know the variance products, $\{\gamma_{lk}^2\}$, of their neighbors, which in general are not available since they require knowledge of the quantities $\{\sigma_{v,l}^2, \text{Tr}(R_{u,l}), \text{Tr}(R_{v,lk}^{(\psi)})\}$. Therefore, we now propose an adaptive combination rule by using data that are available to the individual nodes. For the ATC algorithm, we first note from (24) and (29) that

$$\begin{aligned} \mathbb{E} \|\psi_{lk,i} - \mathbf{w}_{l,i-1}\|^2 &\approx \mu_l^2 \sigma_{v,l}^2 \text{Tr}(R_{u,l}) + \text{Tr}(R_{v,lk}^{(\psi)}) \\ &= \gamma_{lk}^2 \end{aligned} \quad (126)$$

for $l \in \mathcal{N}_k \setminus \{k\}$. Since the algorithm converges in the mean and mean-square senses under Assumption 4.1, all the estimates $\{\mathbf{w}_{k,i}\}$ tend close to w^o as $i \rightarrow \infty$. This allows us to estimate γ_{lk}^2 for node k by using instantaneous realizations of $\|\psi_{lk,i} - \mathbf{w}_{k,i-1}\|^2$, where we replace $\mathbf{w}_{l,i-1}$ by $\mathbf{w}_{k,i-1}$. Similarly, for node k itself, we can use realizations of $\|\psi_{k,i} - \mathbf{w}_{k,i-1}\|^2$ to estimate γ_{kk}^2 . To unify the notation, we define $\psi_{kk,i} \triangleq \psi_{k,i}$. Let $\hat{\gamma}_{lk}^2(i)$ denote an estimator for γ_{lk}^2 that is computed by node k at time i . Then, one way to evaluate $\hat{\gamma}_{lk}^2(i)$ is through the recursion:

$$\boxed{\hat{\gamma}_{lk}^2(i) = (1 - \nu_k) \hat{\gamma}_{lk}^2(i-1) + \nu_k \|\psi_{lk,i} - \mathbf{w}_{k,i-1}\|^2} \quad (127)$$

for $l \in \mathcal{N}_k$, where $\nu_k \in (0, 1)$ is a forgetting factor that is usually close to one. In this way, we arrive at the adaptive combination rule:

$$\boxed{\mathbf{a}_{lk}(i) = \begin{cases} \frac{[\hat{\gamma}_{lk}^2(i)]^{-1}}{\sum_{m \in \mathcal{N}_k} [\hat{\gamma}_{mk}^2(i)]^{-1}}, & \text{if } l \in \mathcal{N}_k \\ 0, & \text{otherwise} \end{cases}} \quad (128)$$

VII. MEAN-SQUARE TRACKING BEHAVIOR

The diffusion strategy (5)–(7) is adaptive in nature. One of the main benefits of adaptation (by using constant step-sizes) is that it endows networks with tracking abilities when the underlying weight vector w^o varies with time. In this section we analyze how well an adaptive network is able to track variations in w^o . To do so, we adopt a random-walk model for w^o that is commonly used in the literature to describe the non-stationarity of the weight vector [32].

Assumption 7.1 (Random-walk model): The weight vector w^o changes according to the model:

$$\mathbf{w}_i^o = \mathbf{w}_{i-1}^o + \boldsymbol{\eta}_i \quad (129)$$

where $\{\mathbf{w}_i^o\}$ has a constant mean w^o for all i , $\{\boldsymbol{\eta}_i\}$ is an i.i.d. random sequence with zero mean and covariance matrix R_η ; the sequence $\{\boldsymbol{\eta}_i\}$ is independent of the initial conditions $\{\mathbf{w}_{-1}^o, \mathbf{w}_{k-1}^o\}$ and of all regression data and noise signals across the network for all time instants. \blacksquare

We now define the error vector at node k as

$$\tilde{\mathbf{w}}_{k,i} \triangleq \mathbf{w}_i^o - \mathbf{w}_{k,i} \quad (130)$$

so that the global error recursion (52) for the network is replaced by

$$\begin{aligned} \tilde{\mathbf{w}}_i &= \mathcal{A}_2^\top (I_{NM} - \mathcal{M}\mathcal{R}'_i) \mathcal{A}_1^\top \tilde{\mathbf{w}}_{i-1} + \mathcal{A}_2^\top (I_{NM} - \mathcal{M}\mathcal{R}'_i) \mathcal{A}_1^\top \boldsymbol{\zeta}_i \\ &\quad - \mathcal{A}_2^\top (I_{NM} - \mathcal{M}\mathcal{R}'_i) \mathbf{v}_{i-1}^{(w)} - \mathcal{A}_2^\top \mathcal{M}\mathbf{z}_i - \mathbf{v}_i^{(\psi)} \end{aligned} \quad (131)$$

where the $NM \times 1$ vector $\boldsymbol{\zeta}_i$ is defined as

$$\boldsymbol{\zeta}_i \triangleq \text{col}\{\boldsymbol{\eta}_i, \dots, \boldsymbol{\eta}_i\} = \mathbb{1}_N \otimes \boldsymbol{\eta}_i \quad (132)$$

A. Convergence Conditions

By Assumptions 2.1 and 7.1, it can be verified that the condition for mean convergence continues to be $\rho(\mathcal{B}) < 1$, where \mathcal{B} is defined in (54). In addition, it can also be verified that the error recursion (131) converges in the mean sense to the same non-zero bias vector g as in (66). From (131) and under Assumption 4.1, we can derive the weighted variance relation:

$$\begin{aligned} \mathbb{E}\|\tilde{\mathbf{w}}_i\|_\sigma^2 &\approx \mathbb{E}\|\tilde{\mathbf{w}}_{i-1}\|_{\mathcal{F}\sigma}^2 + \mathbb{E}\|\mathcal{A}_2^\top (I_{NM} - \mathcal{M}\mathcal{R}'_i) \mathcal{A}_1^\top \boldsymbol{\zeta}_i\|_\sigma^2 \\ &\quad - 2\text{Re}\{\mathbb{E}[\mathbf{z}_i^* \mathcal{M} \mathcal{A}_2 \Sigma \mathcal{A}_2^\top (I_{NM} - \mathcal{M}\mathcal{R}'_i) \mathcal{A}_1^\top \tilde{\mathbf{w}}_{i-1}]\} \\ &\quad + \mathbb{E}\|\mathcal{A}_2^\top (I_{NM} - \mathcal{M}\mathcal{R}'_i) \mathbf{v}_{i-1}^{(w)}\|_\sigma^2 \\ &\quad + \mathbb{E}\|\mathcal{A}_2^\top \mathcal{M}\mathbf{z}_i\|_\sigma^2 + \mathbb{E}\|\mathbf{v}_i^{(\psi)}\|_\sigma^2 \end{aligned} \quad (133)$$

where \mathcal{F} is given in (73). If the step-sizes are sufficiently small, then we can assume that the network continues to be mean-square stable.

B. Steady-State Performance

The steady-state performance is affected by the non-stationarity of w^o . From Assumption 4.1, at steady-state, expression (133) becomes

$$\lim_{i \rightarrow \infty} \mathbb{E} \|\tilde{w}_i\|_{\Omega}^2 \approx [\text{vec}(\mathcal{A}_2^T \mathcal{M} \mathcal{C}^T \mathcal{S} \mathcal{C} \mathcal{M} \mathcal{A}_2 + \mathcal{A}_2^T \mathcal{A}_1^T \mathcal{R}_\zeta \mathcal{A}_1 \mathcal{A}_2 + \mathcal{R}_v + \mathcal{Y} + \mathcal{Y}^*)]^* (I_{N^2 M^2} - \mathcal{F})^{-1} \text{vec}(\Omega) \quad (134)$$

where \mathcal{S} is given in (76), R_v in (82), \mathcal{Y} in (83), \mathcal{F} in (73), and \mathcal{R}_ζ is the covariance matrix of ζ_i :

$$\mathcal{R}_\zeta \triangleq \mathbb{E} \zeta_i \zeta_i^* = (\mathbb{1}_N \mathbb{1}_N^T) \otimes R_\eta \quad (135)$$

By (8), (22), and (135), we get

$$\begin{aligned} \mathcal{A}_2^T \mathcal{A}_1^T \mathcal{R}_\zeta \mathcal{A}_1 \mathcal{A}_2 &= (\mathcal{A}_2^T \mathcal{A}_1^T \mathbb{1}_N \mathbb{1}_N^T \mathcal{A}_1 \mathcal{A}_2) \otimes R_\eta \\ &= (\mathbb{1}_N \mathbb{1}_N^T) \otimes R_\eta \\ &= \mathcal{R}_\zeta \end{aligned} \quad (136)$$

Then, following the same argument that led to (88), we find that the network MSD is now given by:

$$\text{MSD}_{\text{trk}} \approx \frac{1}{N} [\text{vec}(\mathcal{A}_2^T \mathcal{M} \mathcal{C}^T \mathcal{S} \mathcal{C} \mathcal{M} \mathcal{A}_2 + \mathcal{R}_\zeta + \mathcal{R}_v + \mathcal{Y} + \mathcal{Y}^*)]^* (I_{N^2 M^2} - \mathcal{F})^{-1} \text{vec}(I_{NM}) \quad (137)$$

Similarly, the network EMSE is given by:

$$\text{EMSE}_{\text{trk}} \approx \frac{1}{N} [\text{vec}(\mathcal{A}_2^T \mathcal{M} \mathcal{C}^T \mathcal{S} \mathcal{C} \mathcal{M} \mathcal{A}_2 + \mathcal{R}_\zeta + \mathcal{R}_v + \mathcal{Y} + \mathcal{Y}^*)]^* (I_{N^2 M^2} - \mathcal{F})^{-1} \text{vec}(\mathcal{R}_u) \quad (138)$$

where \mathcal{R}_u is defined in (89). Observe that the main difference relative to (88) and (90) is the addition of the term \mathcal{R}_ζ . Therefore, all the results that were derived in the earlier section, such as (95) and (96), continue to hold by adding \mathcal{R}_ζ . In particular, if Assumptions 4.1 and 5.2 are adopted, expressions (137) and (138) can be approximated as

$$\text{MSD}_{\text{trk}} \approx \frac{1}{N} [\text{vec}(\mathcal{A}_2^T \mathcal{M} \mathcal{S} \mathcal{M} \mathcal{A}_2 + \mathcal{R}_\zeta + \mathcal{R}_v)]^* (I_{N^2 M^2} - \mathcal{F})^{-1} \text{vec}(I_{NM}) \quad (139)$$

and

$$\text{EMSE}_{\text{trk}} \approx \frac{1}{N} [\text{vec}(\mathcal{A}_2^T \mathcal{M} \mathcal{S} \mathcal{M} \mathcal{A}_2 + \mathcal{R}_\zeta + \mathcal{R}_v)]^* (I_{N^2 M^2} - \mathcal{F})^{-1} \text{vec}(\mathcal{R}_u) \quad (140)$$

where \mathcal{R}_v is now given in (93).

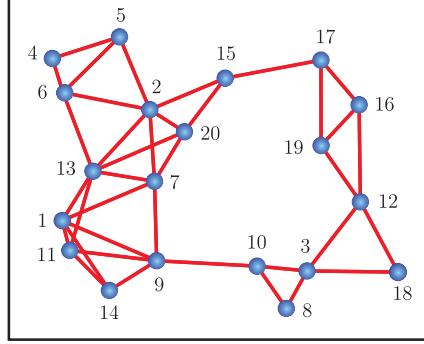


Fig. 2. A network topology with $N = 20$ nodes.

VIII. SIMULATION RESULTS

We simulate two scenarios: noisy information exchanges and non-stationary environments. We consider a connected network with $N = 20$ nodes. The network topology is shown in Fig. 2.

A. Imperfect Information Exchange

The unknown complex parameter w^o of length $M = 2$ is randomly generated; its value is $[0.3750 + j2.0834, 0.7174 + j1.4123]$. We adopt uniform step-sizes, $\{\mu_k = 0.01\}$, and uniformly white Gaussian regression data with covariance matrices $\{R_{u,k} = \sigma_{u,k}^2 I_M\}$, where $\{\sigma_{u,k}^2\}$ are shown in Fig. 3a. The variances of the model noises, $\{\sigma_{v,k}^2\}$, are randomly generated and shown in Fig. 3b. We also use white Gaussian link noise signals such that $R_{v,lk}^{(w)} = \sigma_{w,lk}^2 I_M$, $R_{v,lk}^{(u)} = \sigma_{u,lk}^2 I_M$, and $R_{v,lk}^{(\psi)} = \sigma_{\psi,lk}^2 I_M$. All link noise variances, $\{\sigma_{w,lk}^2, \sigma_{v,lk}^2, \sigma_{u,lk}^2, \sigma_{\psi,lk}^2\}$, are randomly generated and illustrated in Fig. 4 from top to bottom. We assign the link number by the following procedure. We denote the link from node l to node k as $\ell_{l,k}$, where $l \neq k$. Then, we collect the links $\{\ell_{l,k}, l \in \mathcal{N}_k \setminus \{k\}\}$ in an ascending order of l in the list \mathcal{L}_k (which is a set with *ordered* elements) for each node k . For example, for node $k = 2$ in Fig. 2, it has 6 links; the ordered links are then collected in $\mathcal{L}_2 \triangleq \{\ell_{5,2}, \ell_{6,2}, \ell_{7,2}, \ell_{13,2}, \ell_{15,2}, \ell_{20,2}\}$. We concatenate $\{\mathcal{L}_k\}$ in an ascending order of k to get the overall list $\mathcal{L} \triangleq \{\mathcal{L}_1, \mathcal{L}_2, \dots, \mathcal{L}_N\}$. Eventually, the m th link in the network is given by the m th element in the list \mathcal{L} .

We examine the simplified CTA and ATC algorithms in (3) and (4), namely, no sharing of data among nodes (i.e., $C = I_N$), under various combination rules: (i) the relative variance rule in (124), (ii) the

Metropolis rule in [39]:

$$\begin{cases} a_{lk} = \frac{1}{\max\{|\mathcal{N}_k|, |\mathcal{N}_l|\}}, & l \in \mathcal{N}_k \setminus \{k\} \\ a_{lk} = 1 - \sum_{l \in \mathcal{N}_k \setminus \{k\}} a_{lk}, & l = k \\ a_{lk} = 0, & l \notin \mathcal{N}_k \end{cases} \quad (141)$$

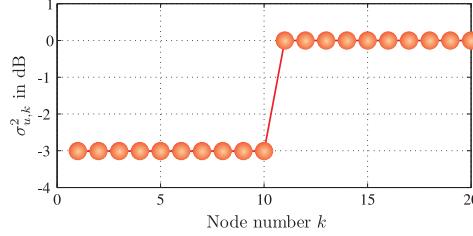
where $|\mathcal{N}_k|$ denotes the degree of node k (including the node itself), (iii) the uniform weighting rule:

$$\begin{cases} a_{lk} = \frac{1}{|\mathcal{N}_k|}, & l \in \mathcal{N}_k \\ a_{lk} = 0, & l \notin \mathcal{N}_k \end{cases} \quad (142)$$

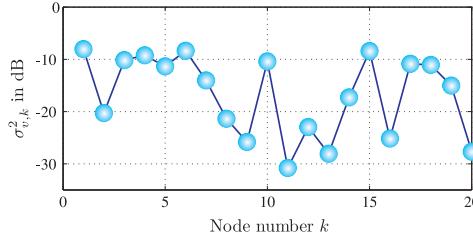
and (iv) the adaptive rule in (128) with $\{\nu_k = 0.05\}$. We plot the network MSD and EMSE learning curves for ATC algorithms in Figs. 5a and 5c by averaging over 50 experiments. For CTA algorithms, we plot their network MSD and EMSE learning curves in Figs. 5b and 5d also by averaging over 50 experiments. Moreover, we also plot their theoretical results (95) and (96) in the same figures. From Fig. 5 we see that the relative variance rule makes diffusion algorithms achieve the lowest MSD and EMSE levels at steady-state, compared to the metropolis and uniform rules as well as the algorithm from [33] (which also requires knowledge of the noise variances). In addition, the adaptive rule attains MSD and EMSE levels that are only slightly larger than those of the relative variance rule, although, as expected, it converges slower due to the additional learning step (127).

B. Non-stationary Scenario

The value for each entry of the complex parameter $w_i^o = \text{col}\{w_{i,1}^o, w_{i,2}^o\}$ is assumed to be changing over time along a circular trajectory in the complex plane, as shown in Fig. 6. The dynamic model for w_i^o is expressed as $w_{i,m}^o = e^{j\omega} w_{i-1,m}^o$, where $m = 1, 2$, $\omega = 2\pi/6000$, and $w_{-1}^o = \text{col}\{1 + j, -1 - j\}$. The covariance matrices $\{R_{u,k}\}$ are randomly generated such that $R_{u,k} \neq R_{u,l}$ when $k \neq l$, but their traces are normalized to be one, i.e., $\text{Tr}(R_{u,k}) = 1$, for all nodes. The variances for the model noises, $\{\sigma_{v,k}^2\}$, are also randomly generated. We examine two different scenarios: the low noise-level case where the average noise variance across the network is -5 dB and the noise variances are shown in Fig. 7a; and the high noise-level case where the average variance is 25 dB and the variances are shown in Fig. 7b. We simulate 3000 iterations and average over 20 experiments in Figs. 6a and 6b for each case. The step-size is 0.01 and uniform across the network. For simplicity, we adopt the simplified ATC algorithm where $C = I_N$, and only use the uniform weighting rule (142). The tracking behavior of the network, denoted

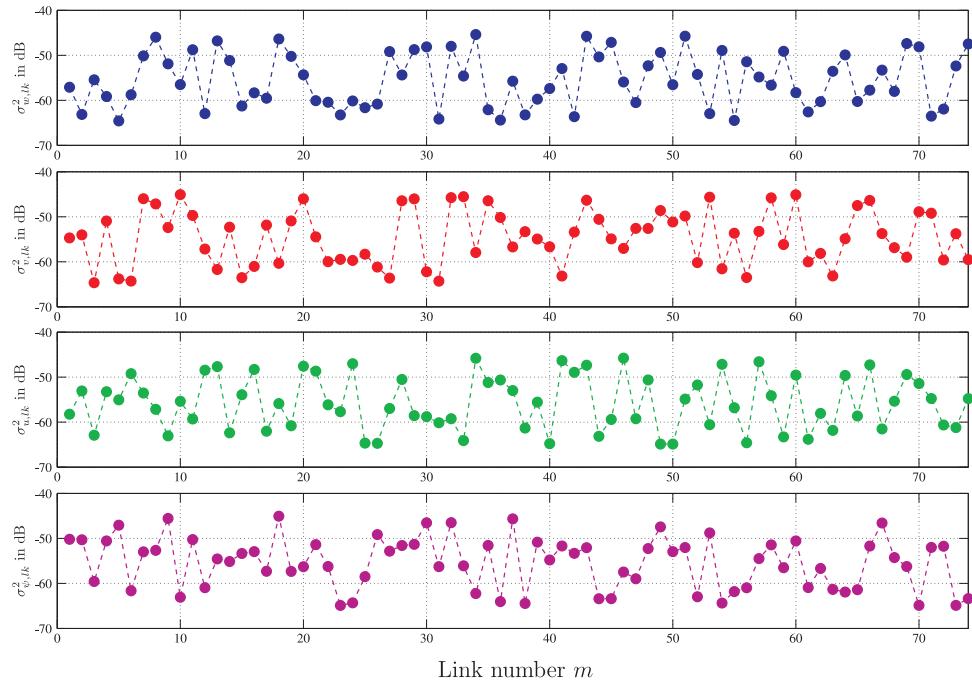


(a) The variance profile of regression data.



(b) The variance profile of measurement noises.

Fig. 3. The variance profiles for regression data and measurement noises.

Fig. 4. The variance profiles for various sources of link noises, including $\{\sigma_w^2, \sigma_v^2, \sigma_u^2, \sigma_\psi^2\}$.

as $\bar{w}_i = \text{col}\{\bar{w}_{i,1}, \bar{w}_{i,2}\}$, is obtained by averaging over all the estimates, $\{w_{k,i}\}$, across the network. Figs. 6a and 6b depict the complex plane; the horizontal axis is the real axis and the vertical axis is the imaginary axis. Therefore, for every time i , each entry of w_i^o or \bar{w}_i represents a point in the plane. When

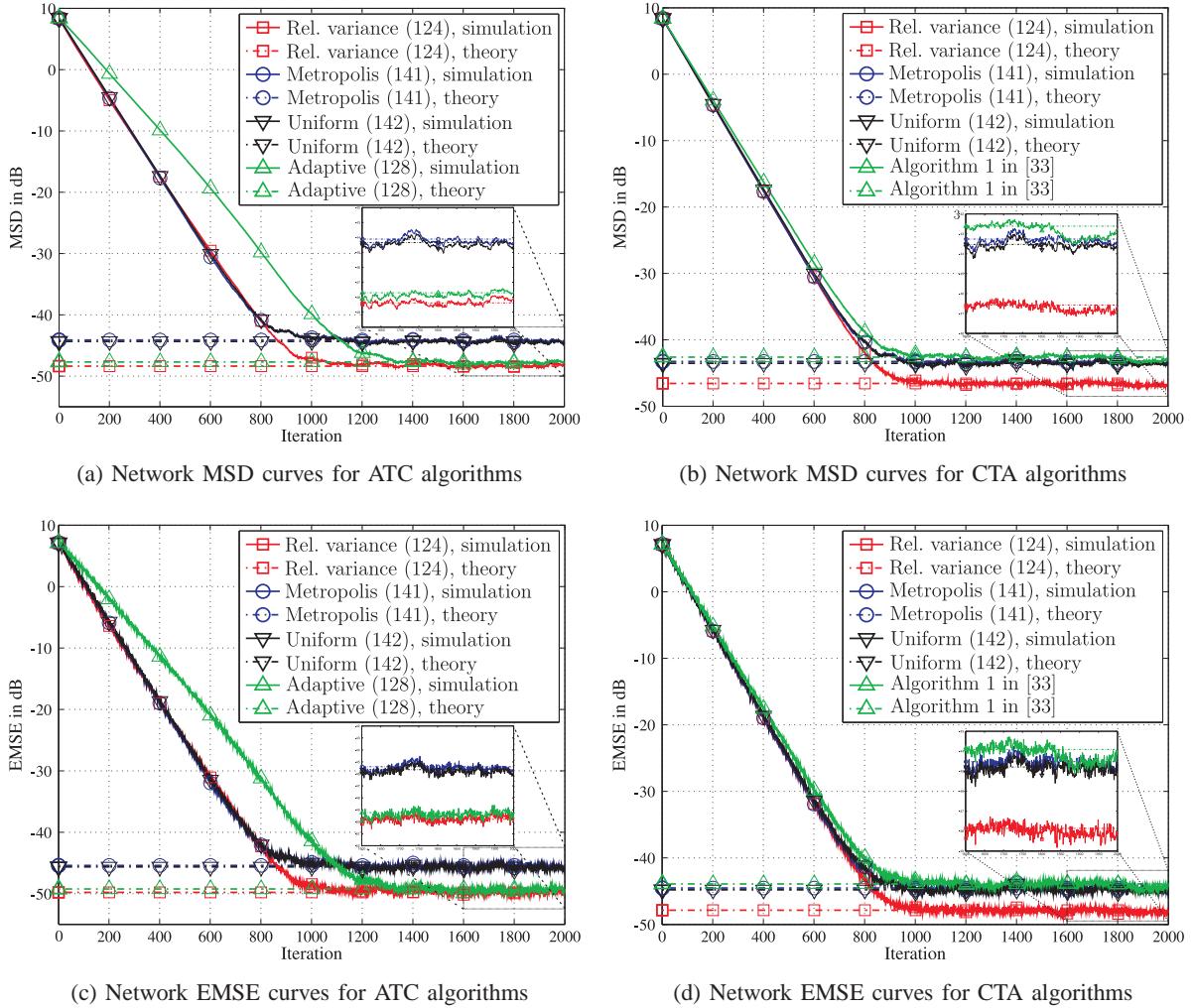
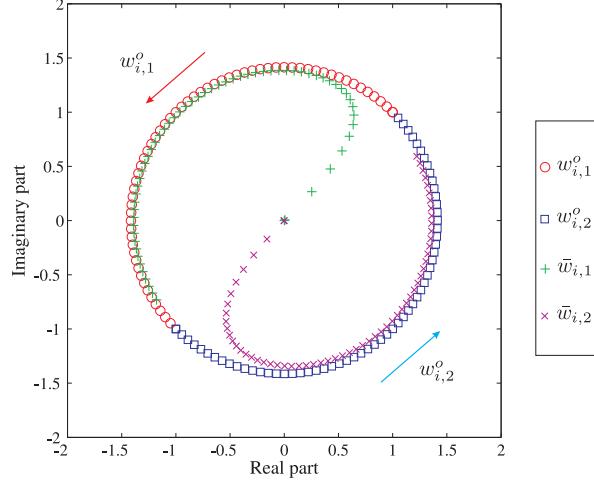


Fig. 5. Simulated network MSD and EMSE curves and theoretical results (95) and (96) for diffusion algorithms with various combination rules under noisy information exchange.

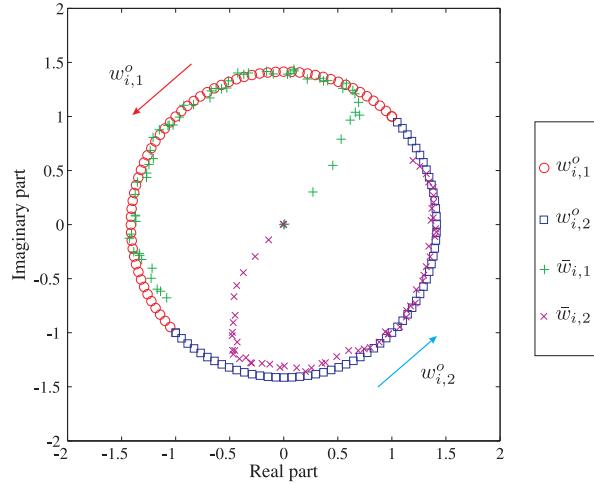
i is increasing, $w_{i,1}^o$ moves along the red trajectory (in \circ), $w_{i,2}^o$ along the blue trajectory (in \square), $\bar{w}_{i,1}$ along the green trajectory (in $+$), and $\bar{w}_{i,2}$ along the magenta trajectory (in \times). From Fig. 6, it can be seen that diffusion algorithms exhibit the tracking ability in both high and low noise-level environments.

IX. CONCLUSIONS

In this work we investigated the performance of diffusion algorithms under several sources of noise during information exchange and under non-stationary environments. We first showed that, on one hand, the link noise over the regression data biases the estimators and deteriorates the conditions for mean and mean-square convergence. On the other hand, diffusion strategies can still stabilize the mean and mean-square convergence of the network with noisy information exchange. We derived analytical expressions



(a) The low noise-level case.



(b) The high noise-level case.

Fig. 6. An adaptive network tracking a parameter vector $w^o \in \mathbb{C}^2$.

for the network MSD and EMSE and used these expressions to motivate the choice of combination weights that help ameliorate the effect of information-exchange noise and improve network performance. We also extended the results to the non-stationary scenario where the unknown parameter w^o is changing over time. Simulation results illustrate the theoretical findings and how well they match with theory.

APPENDIX A

STABILITY OF $\mathcal{A}_2^T (I_{NM} - \mathcal{M}\mathcal{R}') \mathcal{A}_1^T$

Following [15], we first define the block maximum norm of a vector.

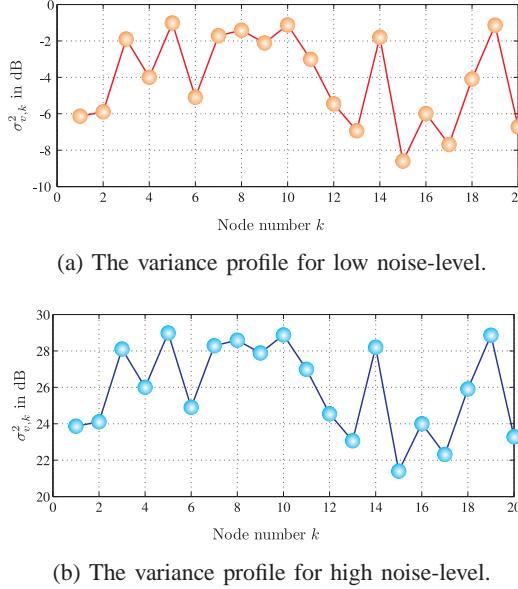


Fig. 7. The noise variance profiles for two cases.

Definition A.1 (Block Maximum Norm): Given a vector $x = \text{col}\{x_1, \dots, x_N\} \in \mathbb{C}^{MN}$ consisting of N blocks $\{x_k \in \mathbb{C}^M, k = 1, \dots, N\}$, the block maximum norm is the real function $\|\cdot\|_{b,\infty} : \mathbb{C}^{MN} \rightarrow \mathbb{R}$, defined as

$$\|x\|_{b,\infty} \triangleq \max_{1 \leq k \leq N} \|x_k\|_2 \quad (143)$$

where $\|\cdot\|_2$ denotes the standard 2-norm on \mathbb{C}^M . ■

Similarly, we define the matrix norm that is induced by the block maximum norm as follows:

Definition A.2 (Block Maximum Matrix Norm): Given a block matrix $\mathcal{A} \in \mathbb{C}^{MN \times MN}$ with block size $M \times M$, then

$$\|\mathcal{A}\|_{b,\infty} \triangleq \max_{x \in \mathbb{C}^{MN} \setminus \{0\}} \frac{\|\mathcal{A}x\|_{b,\infty}}{\|x\|_{b,\infty}} \quad (144)$$

denotes the induced block maximum (matrix) norm on $\mathbb{C}^{MN \times MN}$. ■

Lemma A.3: The block maximum matrix norm is block unitary invariant, i.e., given a block diagonal unitary matrix $\mathcal{U} \triangleq \text{diag}\{U_1, \dots, U_N\} \in \mathbb{C}^{MN \times MN}$ consisting of N unitary blocks $\{U_k \in \mathbb{C}^{M \times M}, k = 1, \dots, N\}$, where $U_k U_k^* = U_k^* U_k = I_M$, for any matrix $\mathcal{A} \in \mathbb{C}^{MN \times MN}$, then

$$\|\mathcal{A}\|_{b,\infty} = \|\mathcal{U}\mathcal{A}\mathcal{U}^*\|_{b,\infty} \quad (145)$$

where $\|\cdot\|_{b,\infty}$ denotes the block maximum matrix norm on $\mathbb{C}^{MN \times MN}$ with block size $M \times M$. ■

Lemma A.4: Let $A \in \mathbb{C}^{N \times N}$ be a right-stochastic matrix. Then, for block size $M \times M$,

$$\|A \otimes I_M\|_{b,\infty} = 1 \quad (146)$$

Proof: From Definition A.2, we get

$$\begin{aligned} \|A \otimes I_M\|_{b,\infty} &= \max_{x \in \mathbb{C}^{MN} \setminus \{0\}} \frac{\max_l \|\sum_{k=1}^N [A]_{lk} x_k\|_2}{\max_k \|x_k\|_2} \\ &\leq \max_{x \in \mathbb{C}^{MN} \setminus \{0\}} \frac{\max_l \sum_{k=1}^N [A]_{lk} \|x_k\|_2}{\max_k \|x_k\|_2} \\ &\leq \max_{x \in \mathbb{C}^{MN} \setminus \{0\}} \frac{\max_l (\sum_{k=1}^N [A]_{lk}) \cdot \max_k \|x_k\|_2}{\max_k \|x_k\|_2} \\ &\leq \max_{x \in \mathbb{C}^{MN} \setminus \{0\}} \frac{\max_l 1 \cdot \max_k \|x_k\|_2}{\max_k \|x_k\|_2} \\ &= 1 \end{aligned} \quad (147)$$

where $x \triangleq \text{col}\{x_1, \dots, x_N\} \in \mathbb{C}^{MN}$ consists of N blocks $\{x_k \in \mathbb{C}^M, k = 1, \dots, N\}$, and $[A]_{lk}$ denotes the (l, k) th entry of A . On the other hand, for any induced matrix norm, say, the block maximum norm, it is always lower bounded by the spectral radius of the matrix [49]:

$$\|A \otimes I_M\|_{b,\infty} \geq \rho(A \otimes I_M) = \rho(A) = 1 \quad (148)$$

Combining (147) and (148) completes the proof. \blacksquare

Lemma A.5: Let $\mathcal{A} \in \mathbb{C}^{NM \times NM}$ be a block diagonal Hermitian matrix with block size $M \times M$. Then the block maximum norm of the matrix \mathcal{A} is equal to its spectral radius, i.e.,

$$\|\mathcal{A}\|_{b,\infty} = \rho(\mathcal{A}) \quad (149)$$

Proof: Denote the k th $M \times M$ submatrix on the diagonal of \mathcal{A} by A_k . Let $A_k = U_k \Lambda_k U_k^*$ be the eigen-decomposition of A_k , where $U_k \in \mathbb{C}^{M \times M}$ is unitary and $\Lambda_k \in \mathbb{R}^{M \times M}$ is diagonal. Define the block unitary matrix $\mathcal{U} \triangleq \text{diag}\{U_1, \dots, U_N\}$ and the diagonal matrix $\Lambda \triangleq \text{diag}\{\Lambda_1, \dots, \Lambda_N\}$. Then,

$\mathcal{A} = \mathcal{U}\Lambda\mathcal{U}^*$. By Lemma A.3, the block maximum norm of \mathcal{A} with block size $M \times M$ is

$$\begin{aligned}
\|\mathcal{A}\|_{b,\infty} &= \|\mathcal{U}\Lambda\mathcal{U}^*\|_{b,\infty} \\
&= \|\Lambda\|_{b,\infty} \\
&= \max_{x \in \mathbb{C}^{MN} \setminus \{0\}} \frac{\max_k \|\Lambda_k x_k\|_2}{\max_k \|x_k\|_2} \\
&\leq \max_{x \in \mathbb{C}^{MN} \setminus \{0\}} \frac{\max_k \|\Lambda_k\|_2 \cdot \|x_k\|_2}{\max_k \|x_k\|_2} \\
&\leq \max_{x \in \mathbb{C}^{MN} \setminus \{0\}} \frac{\max_k \|\Lambda_k\|_2 \cdot \max_k \|x_k\|_2}{\max_k \|x_k\|_2} \\
&= \max_k \|\Lambda_k\|_2 \\
&= \rho(\mathcal{A})
\end{aligned} \tag{150}$$

where we used the fact that the induced 2-norm is identical to the spectral radius for Hermitian matrices [49]. On the other hand, any matrix norm is lower bounded by the spectral radius [49], i.e.,

$$\rho(\mathcal{A}) \leq \|\mathcal{A}\|_{b,\infty} \tag{151}$$

Combining (150) and (151) completes the proof. \blacksquare

Now we show that the matrix $\mathcal{A}_2^T(I_{NM} - \mathcal{M}\mathcal{R}')\mathcal{A}_1^T$ is stable if $I_{NM} - \mathcal{M}\mathcal{R}'$ is stable. For any induced matrix norm, say, the block maximum norm with block size $M \times M$, we have [49]

$$\begin{aligned}
\rho\left(\mathcal{A}_2^T(I_{NM} - \mathcal{M}\mathcal{R}')\mathcal{A}_1^T\right) &\leq \|\mathcal{A}_2^T(I_{NM} - \mathcal{M}\mathcal{R}')\mathcal{A}_1^T\|_{b,\infty} \\
&\leq \|\mathcal{A}_2^T\|_{b,\infty} \cdot \|I_{NM} - \mathcal{M}\mathcal{R}'\|_{b,\infty} \cdot \|\mathcal{A}_1^T\|_{b,\infty} \\
&= \|I_{NM} - \mathcal{M}\mathcal{R}'\|_{b,\infty}
\end{aligned} \tag{152}$$

where, from (8) and (22), \mathcal{A}_1^T and \mathcal{A}_2^T satisfy Lemma A.4. By (17) and (55), it is straightforward to see that $I_{NM} - \mathcal{M}\mathcal{R}'$ is block diagonal with block size $M \times M$. Then, by Lemma A.5, expression (152) can be further expressed as

$$\rho\left(\mathcal{A}_2^T(I_{NM} - \mathcal{M}\mathcal{R}')\mathcal{A}_1^T\right) \leq \rho(I_{NM} - \mathcal{M}\mathcal{R}') \tag{153}$$

which completes the proof.¹

¹This statement fixes the argument that appeared in Appendix I of [11] and Lemma 2 of [12]. Since the matrix \mathcal{X} in Appendix I of [11] and the matrix \mathcal{M} in Lemma 2 of [12] are block diagonal, the $\|\cdot\|_\rho$ norm used in these references should simply be replaced by the $\|\cdot\|_{b,\infty}$ norm used here and as already done in [15].

APPENDIX B

PROOF OF EXPRESSION (75)

Let us denote the (l, k) th submatrix of R_z by $R_{z,lk} \in \mathbb{C}^{M \times M}$. By Assumptions 2.1 and expression (44), $R_{z,lk}$ can be evaluated as

$$\begin{aligned} R_{z,lk} &= \mathbb{E} \mathbf{z}_{l,i} \mathbf{z}_{k,i}^* \\ &= \sum_{m \in \mathcal{N}_l} \sum_{n \in \mathcal{N}_k} c_{ml} c_{nk} \underbrace{\mathbb{E} (\mathbf{u}_{ml,i}^* \mathbf{v}_{ml}(i) \mathbf{v}_{nk}^*(i) \mathbf{u}_{nk,i})}_{\triangleq R_{ml,nk}} \end{aligned} \quad (154)$$

where, by expressions (26) and (38),

$$R_{ml,nk} = \mathbb{E} \left(\mathbf{u}_{m,i} + \mathbf{v}_{ml,i}^{(u)} \right)^* \left(\mathbf{v}_m(i) + \mathbf{v}_{ml}^{(d)}(i) - \mathbf{v}_{ml,i}^{(u)} w^o \right) \left(\mathbf{v}_n(i) + \mathbf{v}_{nk}^{(d)}(i) - \mathbf{v}_{nk,i}^{(u)} w^o \right)^* \left(\mathbf{u}_{n,i} + \mathbf{v}_{nk,i}^{(u)} \right) \quad (155)$$

When $m \neq n$, expression (155) reduces to

$$R_{ml,nk} = R_{v,ml}^{(u)} w^o w^{o*} R_{v,nk}^{(u)} \quad (156)$$

When $m = n$, expression (155) becomes

$$\begin{aligned} R_{ml,nk} &= \mathbb{E} \left(\mathbf{v}_m(i) + \mathbf{v}_{ml}^{(d)}(i) - \mathbf{v}_{ml,i}^{(u)} w^o \right) \left(\mathbf{v}_m(i) + \mathbf{v}_{mk}^{(d)}(i) - \mathbf{v}_{mk,i}^{(u)} w^o \right)^* \left(\mathbf{u}_{m,i} + \mathbf{v}_{ml,i}^{(u)} \right)^* \left(\mathbf{u}_{m,i} + \mathbf{v}_{mk,i}^{(u)} \right) \\ &= \mathbb{E} \left(\mathbf{v}_m(i) + \mathbf{v}_{ml}^{(d)}(i) \right) \left(\mathbf{v}_m(i) + \mathbf{v}_{mk}^{(d)}(i) \right)^* \mathbb{E} \left(\mathbf{u}_{m,i} + \mathbf{v}_{ml,i}^{(u)} \right)^* \left(\mathbf{u}_{m,i} + \mathbf{v}_{mk,i}^{(u)} \right) \\ &\quad + \mathbb{E} \mathbf{v}_{ml,i}^{(u)} w^o w^{o*} \mathbf{v}_{mk,i}^{(u)*} \left(\mathbf{u}_{m,i} + \mathbf{v}_{ml,i}^{(u)} \right)^* \left(\mathbf{u}_{m,i} + \mathbf{v}_{mk,i}^{(u)} \right) \\ &= (\sigma_{v,m}^2 + \delta_{lk} \sigma_{v,ml}^2) \left(R_{u,m} + \delta_{lk} R_{v,ml}^{(u)} \right) + \delta_{lk} w^{o*} R_{v,ml}^{(u)} w^o R_{u,m} + R_{v,ml}^{(u)} w^o w^{o*} R_{v,mk}^{(u)} \\ &\quad + \delta_{lk} \left(\mathbb{E} \mathbf{v}_{ml,i}^{(u)*} \mathbf{v}_{ml,i}^{(u)} w^o w^{o*} \mathbf{v}_{ml,i}^{(u)*} \mathbf{v}_{ml,i}^{(u)} - R_{v,ml}^{(u)} w^o w^{o*} R_{v,ml}^{(u)} \right) \end{aligned} \quad (157)$$

where δ_{lk} denotes the Kronecker delta function. Evaluating the last term on RHS of (157) requires knowledge of the excess kurtosis of $\mathbf{v}_{ml,i}^{(u)}$, which is generally not available. In order to proceed, we invoke a separation principle to approximate it as

$$\mathbb{E} \mathbf{v}_{ml,i}^{(u)*} \mathbf{v}_{ml,i}^{(u)} w^o w^{o*} \mathbf{v}_{ml,i}^{(u)*} \mathbf{v}_{ml,i}^{(u)} \approx R_{v,ml}^{(u)} w^o w^{o*} R_{v,ml}^{(u)} \quad (158)$$

Substituting (158) into (157) leads to

$$\begin{aligned} R_{ml,nk} &\approx (\sigma_{v,m}^2 + \delta_{lk} \sigma_{v,ml}^2) \left(R_{u,m} + \delta_{lk} R_{v,ml}^{(u)} \right) + \delta_{lk} \left(w^{o*} R_{v,ml}^{(u)} w^o \right) R_{u,m} + R_{v,ml}^{(u)} w^o w^{o*} R_{v,mk}^{(u)} \\ &= \sigma_{v,m}^2 R_{u,m} + R_{v,ml}^{(u)} w^o w^{o*} R_{v,mk}^{(u)} + \delta_{lk} \left[(\sigma_{v,ml}^2 + w^{o*} R_{v,ml}^{(u)} w^o) R_{u,m} + (\sigma_{v,m}^2 + \sigma_{v,ml}^2) R_{v,ml}^{(u)} \right] \end{aligned} \quad (159)$$

From (156) and (159), we get

$$\begin{aligned} R_{ml,nk} \approx & R_{v,ml}^{(u)} w^o w^{o*} R_{v,mk}^{(u)} + \delta_{mn} \sigma_{v,m}^2 R_{u,m} \\ & + \delta_{mn} \delta_{lk} \left[(\sigma_{v,ml}^2 + w^{o*} R_{v,ml}^{(u)} w^o) R_{u,m} + (\sigma_{v,m}^2 + \sigma_{v,ml}^2) R_{v,ml}^{(u)} \right] \end{aligned} \quad (160)$$

Substituting (160) into (154), we obtain

$$\begin{aligned} R_{z,lk} \approx & \left(\sum_{m \in \mathcal{N}_l} c_{ml} R_{v,ml}^{(u)} w^o \right) \left(\sum_{n \in \mathcal{N}_k} c_{nk} R_{v,mk}^{(u)} w^o \right)^* + \sum_{m \in \mathcal{N}_l} \sum_{n \in \mathcal{N}_k} c_{ml} c_{nk} \delta_{mn} \sigma_{v,m}^2 R_{u,m} \\ & + \delta_{lk} \sum_{m \in \mathcal{N}_l} c_{ml}^2 \left[(\sigma_{v,ml}^2 + w^{o*} R_{v,ml}^{(u)} w^o) R_{u,m} + (\sigma_{v,m}^2 + \sigma_{v,ml}^2) R_{v,ml}^{(u)} \right] \end{aligned} \quad (161)$$

From (58)–(59) and (76)–(78), we arrive at expression (75).

REFERENCES

- [1] X. Zhao and A. H. Sayed, “Combination weights for diffusion strategies with imperfect information exchange,” in *Proc. IEEE Int. Conf. Commun. (ICC)*, Ottawa, Canada, June 2012.
- [2] J. Tsitsiklis and M. Athans, “Convergence and asymptotic agreement in distributed decision problems,” *IEEE Trans. Autom. Control*, vol. 29, no. 1, pp. 42–50, Jan. 1984.
- [3] J. Tsitsiklis, D. Bertsekas, and M. Athans, “Distributed asynchronous deterministic and stochastic gradient optimization algorithms,” *IEEE Trans. Autom. Control*, vol. 31, no. 9, pp. 803–812, Sept. 1986.
- [4] D. P. Bertsekas, “A new class of incremental gradient methods for least squares problems,” *SIAM J. Optim.*, vol. 7, no. 4, pp. 913–926, 1997.
- [5] A. Nedic and D. P. Bertsekas, “Incremental subgradient methods for nondifferentiable optimization,” *SIAM J. Optim.*, vol. 12, no. 1, pp. 109–138, 2001.
- [6] M. G. Rabbat and R. D. Nowak, “Quantized incremental algorithms for distributed optimization,” *IEEE J. Sel. Areas Commun.*, vol. 23, no. 4, pp. 798–808, Apr. 2005.
- [7] C. G. Lopes and A. H. Sayed, “Incremental adaptive strategies over distributed networks,” *IEEE Trans. Signal Process.*, vol. 48, no. 8, pp. 223–229, Aug. 2007.
- [8] C. G. Lopes and A. H. Sayed, “Distributed processing over adaptive networks,” in *Proc. Adapt. Sensor Array Process. Workshop (ASAP)*, MIT Lincoln Laboratory, MA, June 2006, pp. 1–5.
- [9] C. G. Lopes and A. H. Sayed, “Diffusion least-mean squares over adaptive networks: Formulation and performance analysis,” *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3122–3136, July 2008.
- [10] F. S. Cattivelli and A. H. Sayed, “Diffusion LMS algorithms with information exchange,” in *Proc. Asilomar Conf. Signals, Systems, Computers*, Pacific Grove, CA, Oct. 2008, pp. 251–255.
- [11] F. S. Cattivelli and A. H. Sayed, “Diffusion LMS strategies for distributed estimation,” *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1035–1048, Mar. 2010.
- [12] F. S. Cattivelli and A. H. Sayed, “Diffusion strategies for distributed Kalman filtering and smoothing,” *IEEE Trans. Autom. Control*, vol. 55, no. 9, pp. 2069–2084, Sept. 2010.
- [13] F. S. Cattivelli, C. G. Lopes, and A. H. Sayed, “Diffusion recursive least-squares for distributed estimation over adaptive networks,” *IEEE Trans. Signal Process.*, vol. 56, no. 5, pp. 1865–1877, May 2008.

- [14] L. Li and J. A. Chambers, "Distributed adaptive estimation based on the APA algorithm over diffusion networks with changing topology," in *Proc. IEEE Workshop Stat. Signal Process. (SSP)*, Cardiff, UK, Aug. / Sept. 2009, pp. 757–760.
- [15] N. Takahashi, I. Yamada, and A. H. Sayed, "Diffusion least-mean squares with adaptive combiners: Formulation and performance analysis," *IEEE Trans. Signal Process.*, vol. 58, no. 9, pp. 4795–4810, Sept. 2010.
- [16] N. Takahashi and I. Yamada, "Link probability control for probabilistic diffusion least-mean squares over resource-constrained networks," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Dallas, TX, Mar. 2010, pp. 3518–3521.
- [17] S. Chouvardas, K. Slavakis, and S. Theodoridis, "Adaptive robust distributed learning in diffusion sensor networks," *IEEE Trans. Signal Process.*, vol. 59, no. 10, pp. 4692–4707, Oct. 2011.
- [18] S-Y. Tu and A. H. Sayed, "Mobile adaptive networks," *IEEE J. Sel. Top. Signal Process.*, vol. 5, no. 4, pp. 649–664, Aug. 2011.
- [19] F. S. Cattivelli and A. H. Sayed, "Modeling bird flight formations using diffusion adaptation," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 2038–2051, May 2011.
- [20] J. Li and A. H. Sayed, "Modeling bee swarming behavior through diffusion adaptation with asymmetric information sharing," *EURASIP J. Advances Signal Process.*, pp. 1–14, doi:10.1186/1687-6180-2012-18, Jan. 2012.
- [21] Z. Towfic, J. Chen, and A. H. Sayed, "Collaborative learning of mixture models using diffusion adaptation," in *Proc. IEEE Int. Workshop Machine Learn. Signal Process. (MLSP)*, Beijing, China, Sept. 2011, pp. 1–6.
- [22] Y. Weng, W. Xiao, and L. Xie, "Diffusion-based EM algorithm for distributed estimation of Gaussian mixtures in wireless sensor networks," *Sensors*, vol. 11, no. 6, pp. 6297–6316, June 2011.
- [23] J. Chen and A. H. Sayed, "Diffusion adaptation strategies for distributed optimization and learning over networks," to appear in *IEEE Trans. Signal Process.*, 2012. [Also available online at <http://arxiv.org/abs/1111.0034> as *arXiv:1111.0034v2 [math.OC]*, Oct. 2011.]
- [24] A. Jadbabaie, J. Lin, and A. S. Morse, "Coordination of groups of mobile autonomous agents using nearest neighbor rules," *IEEE Trans. Autom. Control*, vol. 48, no. 6, pp. 988–1001, June 2003.
- [25] J. A. Fax and R. M. Murray, "Information flow and cooperative control of vehicle formations," *IEEE Trans. Autom. Control*, vol. 49, no. 9, pp. 1465–1476, Sept. 2004.
- [26] R. Olfati-Saber, "Flocking for multi-agent dynamic systems: Algorithms and theory," *IEEE Trans. Autom. Control*, vol. 51, no. 3, pp. 401–420, Mar. 2006.
- [27] R. Olfati-Saber and R. M. Murray, "Consensus problems in networks of agents with switching topology and time-delays," *IEEE Trans. Autom. Control*, vol. 49, no. 9, pp. 1520–1533, Sept. 2004.
- [28] S. Barbarossa and G. Scutari, "Bio-inspired sensor network design," *IEEE Signal Process. Mag.*, vol. 24, no. 3, pp. 26–35, May 2007.
- [29] A. Nedic and A. Ozdaglar, "Distributed subgradient methods for multi-agent optimization," *IEEE Trans. Signal Process.*, vol. 54, no. 1, pp. 48–61, Jan. 2009.
- [30] A. G. Dimakis, S. Kar, J. M. F. Moura, M. G. Rabbat, and A. Scaglione, "Gossip algorithms for distributed signal processing," *Proc. IEEE*, vol. 98, no. 11, pp. 1847–1864, Nov. 2010.
- [31] S. Kar and J. M. F. Moura, "Convergence rate analysis of distributed gossip (linear parameter) estimation: Fundamental limits and tradeoffs," *IEEE J. Sel. Top. Signal Process.*, vol. 5, no. 4, pp. 674–690, Aug. 2011.
- [32] A. H. Sayed, *Adaptive Filters*, Wiley, NJ, 2008.

- [33] R. Abdolee and B. Champagne, "Diffusion LMS algorithms for sensor networks over non-ideal inter-sensor wireless channels," in *Proc. IEEE Int. Conf. Dist. Comput. Sensor Systems (DCOSS)*, Barcelona, Spain, June 2011, pp. 1–6.
- [34] A. Khalili, M. A. Tinati, A. Rastegarnia, and J. A. Chambers, "Transient analysis of diffusion least-mean squares adaptive networks with noisy channels," *Int. J. Adapt. Control Signal Process.*, Sept. 2011, doi: 10.1002/acs.1279.
- [35] A. Khalili, M. A. Tinati, A. Rastegarnia, and J. A. Chambers, "Steady-state analysis of diffusion LMS adaptive networks with noisy links," *IEEE Trans. Signal Process.*, vol. 60, no. 2, pp. 974–979, Feb. 2012.
- [36] S-Y. Tu and A. H. Sayed, "Adaptive networks with noisy links," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Houston, TX, Dec. 2011, pp. 1–5.
- [37] S. Kar and J. M. F. Moura, "Distributed consensus algorithms in sensor networks: Link failures and channel noise," *IEEE Trans. Signal Process.*, vol. 57, no. 1, pp. 355–369, Jan. 2009.
- [38] G. Mateos, I. D. Schizas, and G. B. Giannakis, "Performance analysis of the consensus-based distributed LMS algorithm," *EURASIP J. Advances Signal Process.*, pp. 1–19, Article ID 981030, doi:10.1155/2009/981030, 2009.
- [39] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equations of state calculations by fast computing machines," *J. Chem. Phys.*, vol. 21, no. 6, pp. 1087–1092, 1953.
- [40] L. Xiao, S. Boyd, and S. Lall, "A scheme for robust distributed sensor fusion based on average consensus," in *Proc. ACM/IEEE Int. Conf. Inform. Process. Sensor Networks (IPSN)*, Los Angeles, CA, Apr. 2005, pp. 63–70.
- [41] L. Xiao and S. Boyd, "Fast linear iterations for distributed averaging," *System Control Lett.*, vol. 53, no. 9, pp. 65–78, Sept. 2004.
- [42] D. Jakovetic, J. Xavier, and J. M. F. Moura, "Weight optimization for consensus algorithms with correlated switching topology," *IEEE Trans. Signal Process.*, vol. 58, no. 7, pp. 3788–3801, July 2010.
- [43] J. Arenas-Garcia, V. Gomez-Verdejo, and A. R. Figueiras-Vidal, "New algorithms for improved adaptive convex combination of LMS transversal filters," *IEEE Trans. Instrum. Meas.*, vol. 54, no. 6, pp. 2239–2249, Dec. 2005.
- [44] D. Mandic, P. Vayanos, C. Boukis, B. Jelfs, S. L. Goh, T. Gautama, and T. Rutkowski, "Collaborative adaptive learning using hybrid filters," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Honolulu, HI, Apr. 2007, pp. 921–924.
- [45] M. T. M. Silva and V. H. Nascimento, "Improving the tracking capability of adaptive filters via convex combination," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3137–3149, July 2008.
- [46] S. S. Kozat and A. C. Singer, "A performance-weighted mixture of LMS filters," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Taipei, Apr. 2009, pp. 3101–3104.
- [47] R. Candido, M. T. M. Silva, and V. H. Nascimento, "Transient and steady-state analysis of the affine combination of two adaptive filters," *IEEE Trans. Signal Process.*, vol. 58, no. 8, pp. 4064–4078, Aug. 2010.
- [48] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge Univ. Press, 2004.
- [49] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge Univ. Press, Cambridge, UK, 1985.
- [50] T. Y. Al-Naffouri and A. H. Sayed, "Transient analysis of data-normalized adaptive filters," *IEEE Trans. Signal Process.*, vol. 51, no. 3, pp. 639–652, Mar. 2003.
- [51] A. J. Laub, *Matrix Analysis for Scientists and Engineers*, SIAM, PA, 2005.
- [52] S-Y. Tu and A. H. Sayed, "Optimal combination rules for adaptation and learning over netwrks," in *Proc. IEEE Int. Workshop Comput. Advances Multi-Sensor Adapt. Process. (CAMSAP)*, San Juan, Puerto Rico, Dec. 2011, pp. 317–320.